Mirror therapy for the treatment of upper limb impairment after stroke: investigation of the feasibility aims of a pilot randomised controlled trial

Beverley Turtle

BSc. (Hons) Occupational Therapy, 2016
BSc. (Hons) Psychology, 2007

This thesis is submitted for the degree of Doctor of Philosophy

October 2020

Faculty of Life and Health Sciences of Ulster University

I confirm that the word count of this thesis is less than 100, 000 words

Table of Contents

List of Figures

List of Tables

Acknowledgements

The completion of this thesis would not have been possible without the support from a number of people. Thank you to my supervisors, Dr Alison Porter-Armstrong and Dr May Stinson for your continuous guidance, advice and support. Thanks to Dr Alison Porter-Armstrong for your expertise, dedication and uplifting enthusiasm. These have been invaluable in the completion of this thesis and I am honoured to have been afforded the opportunity to learn from you. Thanks to Dr May Stinson, who provided a voice of reason and thoughtful guidance throughout. I appreciate your patience and being available to answer my queries.

Thank you to occupational therapists Patricia McIlwaine, Lourene Abbi and Fiona Morrow in the Northern Health and Social Care Trust who were instrumental in the development and implementation of the mirror therapy pilot trial and provided expert guidance on the scoring of outcome assessments and upper limb treatment following stroke. Thank you to Nicola Gallagher, a researcher on the trial who collected all data and provided support and friendship. Additional thanks to Patricia McIlwaine who supported the focus group study in her department, Jennifer Trainor and Lourene Abbi who gave up their time to assist with the focus group study. Special thanks to Leona Robinson who moderated the focus group and provided additional insight for the analysis.

Special thanks to all the participants who generously contributed to these studies in various ways and shared their experiences. This work would not have been possible without their involvement.

My special thanks to Leslie Gourley who sourced and made test items for the graded Wolf Motor Function Test, and to all the technicians within the School of Life and Health Sciences for answering my queries and requests. Thanks also to Dr Ian Bradbury who provided statistical advice and answered my queries.

I gratefully acknowledge the Department for the Economy for funding my PhD and to the United Kingdom Occupational Therapy Research Foundation who funded the pilot randomised controlled trial of mirror therapy.

Thank you to my fellow colleagues and researchers within the Centre for Health and Rehabilitation Technologies, who always provided a calming, listening ear and a supportive environment.

To my close friends and family thank you for putting up with me these last few years and being my cheerleaders. To my mum who has always believed I can achieve anything I set my mind to and wholeheartedly supporting me in all my ventures. Finally, to my husband Richard who has been steadfast in his support and patience, thank you for all your unheeded advice.

Dissemination

Presentations

Porter-Armstrong, A., Stinson, M.D., Turtle, B., Gallagher, N., McIlwaine, P., Abbi, L. and Morrow, F. (2018) Use of mirror box therapy and the graded Wolf Motor Function Test in upper limb rehabilitation: a practical workshop. Royal College of Occupational Therapists, Annual Conference, Belfast, June 2018. [Oral presentation]

Turtle, B., Porter-Armstrong, A. and Stinson, M.D. (2018) Reliability and responsiveness of outcome measures in a stroke clinical trial. Royal College of Occupational Therapists, Annual Conference, Belfast, UK. [Oral presentation] **Awarded Early Career Researcher Award 2018**.

Turtle, B., Porter-Armstrong, A. and Stinson, M.D. (2019) Using the EQ-5D-5L and Canadian Occupational Performance Measure in stroke rehabilitation: responsiveness and interpretation. Northern Ireland Stroke Conference. Belfast, UK. [Oral presentation]

Turtle, B., Porter-Armstrong, A. and Stinson, M.D. (November 2020) Experiences of mirror therapy in the sub-acute phase of stroke. European Stroke Organisation and World Stroke Organisation Conference. Vienna, Austria. [Poster presentation]

Publications:

Turtle, B., Porter-Armstrong, A. and Stinson, M. (2019) A systematic review of the application and psychometric properties of the graded Wolf Motor Function Test. *British Journal of Occupational Therapy*, 83(5), 285-296.

Turtle, B., Porter-Armstrong, A. and Stinson, M. (2020) The reliability of the graded Wolf Motor Function Test for stroke. *British Journal of Occupational Therapy*, In Press.

Abstract

The aim of this thesis was to assess the feasibility of the outcome measures included in a pilot randomised controlled trial of mirror therapy in upper limb rehabilitation and to explore the acceptability of mirror therapy within three months of stroke onset.

A systematic literature review investigated the contextual application and psychometric properties of the graded Wolf Motor Function Test, an upper limb assessment tool and indicated psychometric evaluation was essential. Assessment of the reliability (n=30) and responsiveness (n=40) of the graded Wolf Motor Function Test was completed, followed by assessment of the responsiveness of the Functional Independence Measure (n=40) and patient-reported outcome measures (n=39) included in the pilot trial. Finally, a focus group study was conducted to explore perceptions of mirror therapy by stroke survivors (n=3).

The graded Wolf Motor Function Test was reliable and responsive, and findings indicated the use of video for scoring may not be required. The Functional Independence Measure and patient-reported outcome measures demonstrated adequate responsiveness. However, in light of the level of attrition found by six-month follow-up, continuing assessments at this time point was not recommended. Themes identified from the focus group indicated the importance of therapist actions to facilitate stroke survivors to self-manage their recovery within and beyond the hospital setting.

This PhD thesis has contributed to the development of a main trial examining the effectiveness of mirror therapy, which is currently underway. Although considered important to stroke survivors, patient-reported outcome measures are not often reported in clinical trials and this is the first study to assess the responsiveness of

those included, in the early stages of stroke. The focus group study provided a novel exploration of stroke survivors perceptions of mirror therapy treatment. This thesis contributes to the choice of outcome measure and treatment by occupational therapists in upper limb stroke rehabilitation.

Abbreviations

ADLs	Activities of daily living
APA	Alison Porter-Armstrong
AUC	Area under the curve
BT	Beverley Turtle
CASP	Critical appraisal skills programme
CI	Confidence interval
COPM	Canadian Occupational Performance Measure
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
DALYs	Disability-adjusted life years
EMM	Estimated marginal mean
ES	Effect size
FAS	Functional ability scale
FIM	Functional Independence Measure
gWMFT	graded Wolf Motor Function Test
HRQOL	Health-related quality of life
ICC	Intraclass correlation coefficient
ICF	International Classification of Functioning, Disability and Health
LR	Leona Robinson
MCID	Minimal clinically important difference
MS	May Stinson
NHSCT	Northern Health and Social Care Trust
NICE	National Institute for Health and Care Excellence
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PROMs	Patient-reported outcome measures
RCT	Randomised controlled trial

ROC	Receiver operating characteristic
RQIA	The Regulation and Quality Improvement Authority
SD	Standard deviation
SE	Standard error
SEM	Standard error of measurement
SF-36	Short Form 36
SRM	Standardised response mean
UK	United Kingdom
VAS	Visual analogue scale
WHO	World Health Organisation
WMFT	Wolf Motor Function Test

Chapter 1

Chapter 1 - Introduction

1.1 Stroke

Stroke, also known as cerebrovascular accident, is a disease brought about by an interruption to the blood supply in the brain and is defined by the World Health Organisation (WHO) as “rapidly developing clinical signs of focal (or global) disturbance of cerebral function, with vascular origin” (Hatano 1976). This is the result of either ischaemic or haemorrhagic injury leading to the damage or death of brain cells (Bartels *et al*. 2016). Ischaemic stroke occurs in just over 80% of strokes (GBD 2016 Stroke Collaborators 2019), and results from the blockage of a blood vessel (Bartels *et al*. 2016). Although occurring to a lesser degree, haemorrhagic stroke is associated with poorer prognosis, with greater incidences of death and disability (GBD 2016 Stroke Collaborators 2019).

Stroke is the second leading cause of death globally and among the primary causes of disability-adjusted life years (DALYs) (GBD 2017 DALYs and HALE Collaborators 2018). It is a major cause of complex adult disability, with over 80.1 million people living with stroke (GBD 2016 Stroke Collaborators 2019). While mortality rates for stroke have declined, the incidence of stroke Is projected to continue to rise due to an increasingly ageing population (Béjot *et al*. 2016). By 2035 the incidence of stroke is projected to increase by 34% in the European Union (Stevens *et al*. 2017). Taken alongside increasing rates of ischemic and haemorrhagic stroke among young adults (20-64 years) (Krishnamurthi *et al*. 2015), there are significant implications for increased DALYs and reduced productivity, as more people survive and live with the long-term consequences.

In the United Kingdom (UK) more than 100,000 strokes occur each year, with 1.2 million people living with the long-term consequences of stroke (Stroke Association 2018). In 2019 more than 38,000 people were on the stroke register in Northern Ireland (Northern Ireland Statistics and Research Agency 2019) and more than half of stroke survivors are reported to be under the age of 75 (British Heart Foundation 2020). Financially, stroke costs the UK approximately £9 billion per year; incorporating health care costs and informal care costs (Saka *et al.* 2009). Approximately 40% of stroke survivors discharged home in England, Wales and Northern Ireland continue to require assistance with completion of activities of daily living (ADLs) (Royal College of Physicians, Clinical Effectiveness and Evaluation Unit on behalf of the Intercollegiate Stroke Working Party 2016). Significant challenges in Northern Ireland stroke care result from the combination of an ageing population living longer and the prevalence of stroke. Work is currently underway to reconfigure stroke services for improved care and outcomes (Department of Health 2019).

Stroke is a heterogenous disease, the impact of which is determined by the area of the brain affected and the size of the lesion (Langhorne *et al*. 2009). Stroke can lead to variable impairments across cognitive, physical and psychosocial domains, leading to difficulties with communication, swallowing, movement control, memory, fatigue and changes in mood (Walker *et al.* 2013). The International Classification of Functioning, Disability and Health (ICF) provides a framework through which the multi-faceted consequences of stroke can be understood, across the domains of body function, activity and participation (WHO 2001). The ICF is a biopsychosocial model of health and includes consideration for how these domains interact with personal and environmental factors to guide person-centred rehabilitation.

1.2 Upper limb

Motor impairment affects 80% of stroke survivors (Langhorne *et al*. 2009), predominantly affecting one side of the body contralateral to the site of the brain lesion, with pronounced consequences for upper limb function (Lawrence *et al*. 2001; Gillen and Nilsen 2016). Upper limb impairment can involve spasticity, poor motor control, paresis and somatosensory deficits, with stroke survivors generally experiencing these in combination (Lang *et al*. 2013). Upper limb function is integral to the completion of ADLs as it impacts on the ability to wash, dress, go to the toilet, cook, clean and drive (Houwink *et al*. 2013). Overall these deficits contribute to a lack of active and coordinated movement that is detrimental to functional use of the upper limb.

Recovery of upper limb function has been linked to the degree of initial severity of the stroke (Coupar *et al*. 2012) and over half of those who present with severe impairments fail to gain functional recovery six months later (Kwakkel *et al*. 2003). In a study of 421 participants, Nakayama *et al*. (1994) found 70% of stroke survivors with mild impairments regained full upper limb function in comparison to 18% of those with severe impairment. With many individuals failing to gain full recovery of function, upper limb impairment is linked to increased levels of anxiety (Morris *et al*. 2013) and reduced life satisfaction (Ostwald *et al*. 2009). Improvements in upper limb motor control and function are central to stroke rehabilitation (Pollock *et al*. 2014).

1.2.1 Neuroplasticity

The brain is capable of reorganisation; to change structure, function and its connections in response to internal and external stimuli (Murphy and Corbett 2009; Cramer *et al*. 2011). Neuroplasticity is evident following stroke as changes take place

in the brain in response to the damage in order to facilitate motor recovery (Murphy and Corbett 2009). Recovery may occur through the recruitment of existing ipsilesional and contralesional latent connections and through the formation of new neuronal connections (Murphy and Corbett 2009). Although how these influence the recovery of movement and function remains unknown (Buma *et al*. 2013), there is evidence that meaningful, repetitive movements have the potential to drive neuroplastic changes (Arya *et al*. 2011). With the optimum period of neuroplasticity believed to occur up to three months following stroke (Krakauer *et al* 2012), it is integral that stroke survivors receive evidence-based upper limb treatments capable of driving these neuroplastic changes.

1.3 Rehabilitation

The aim of stroke rehabilitation is to promote optimal functioning, participation and quality of life with specialist multidisciplinary involvement (National Institute for Health and Care Excellence (NICE) 2013). Rehabilitation is integral to supporting individuals to return home, to live as independently as possible and engage in meaningful aspects of societal life. Stroke rehabilitation predominantly takes place within the first months of stroke (Intercollegiate Stroke Working Party 2016). There is strong evidence that stroke care delivered by an organised multidisciplinary team in an inpatient setting leads to reduced mortality rates, with stroke survivors more likely to return home and have reduced levels of disability (Langhorne *et al*. 2020). The multidisciplinary team comprises individuals across medical, nursing, allied health and social work fields. Additional features of an inpatient stroke unit include routine training of staff and involvement of caregivers in the rehabilitation process (Langhorne *et al*. 2020).

Stroke guidelines advocate that rehabilitation must be individualised and person-centred, directed at supporting the specific goals and preferences of stroke survivors

(Intercollegiate Stroke Working Party 2016). This incorporates a cyclical process of assessment, goal setting, delivery of interventions to meet goals and re-assessment, essential for the evaluation of patient progress (Langhorne *et al*. 2011). In line with interventions targeted at patient goals, more therapy input is recommended with meaningful, task specific interventions for improved outcomes (Langhorne *et al*. 2011). National Institute for Health and Care Excellence guidelines recommend at least 45 minutes of each required therapy at least five days per week (NICE 2013). Stroke outcomes remain varied across individuals and considering the poor outcomes for those with severe upper limb function, rehabilitation plays an essential role in enhancing recovery (Winstein *et al*. 2016).

1.3.1 Upper limb treatment

Improvement in upper limb function is often a key goal of stroke survivors (Lang *et al*. 2013) and increased research in this area has been identified as a priority by stroke survivors, caregivers and health professionals (Pollock *et al*. 2012). Occupational therapists and physiotherapists are mainly responsible for treatment of the upper limb following stroke (Pollock *et al*. 2014). However, treatment of the upper limb poses a challenge in regard to the complexity involved in motor control of the hand and arm (Lang *et al*. 2013) and the heterogenous outcomes across stroke survivors (Langhorne *et al*. 2011). To meet the complexity involved in movement of the hand and arm, interventions may be delivered individually or in combination, targeted at reducing impairments and improving function in line with patient goals (Pollock *et al*. 2014).

A Cochrane overview of systematic reviews found no high-quality evidence for the range of upper limb interventions in use (Pollock *et al*. 2014). While UK stroke guidelines recommend constraint-induced movement therapy, mental imagery, virtual

reality, repetitive task training and mirror therapy for treatment of the arm alongside conventional therapies (Intercollegiate Stroke Working Party 2016), the evidence base for all is lacking, and further research is needed (Pollock *et al*. 2014). This factor is compounded by quantitative and qualitative studies reporting reduced focus on the upper limb during rehabilitation with clinician focus aimed at improving mobility to meet discharge demands (Barker and Brauer 2005; Lang *et al*. 2009; Meadmore *et al*. 2019).

1.3.1.1 Mirror therapy

Mirror therapy provides an alternative approach to upper limb treatment that does not require a minimum level of voluntary movement. Based on visual feedback, mirror therapy can be applied across all levels of stroke severity. It encourages active involvement from the stroke survivor and has the potential to allow individuals to deliver the treatment for themselves (Thieme *et al.* 2018). Due to the potential for minimal therapist input and equipment requirements, mirror therapy provides a cost-effective approach (Yavuzer *et al.* 2008). Delivering effective and efficient stroke care is essential with an ageing population and increasing financial constraints on the National Health Service (Timmins 2013).

What is mirror therapy?

A mirror is placed in the mid-sagittal plane between the two limbs, with the paretic limb positioned behind the mirror and the non-paretic limb positioned facing the mirror (Ramachandran and Altschuler 2009). The reflection of the non-paretic limb then appears in the same position of the paretic limb. As a result, movements of the non-paretic limb in the mirror can be attributed to the paretic limb.

This process was originally devised to reduce phantom limb pain following amputation (Ramachandran and Rogers-Ramachandran 1996). Following limb removal individuals may continue to feel the presence of the amputated limb and experience debilitating pain. The use of visual feedback provided by the mirror led to a reduction in pain reported by some participants (Ramachandran and Rogers-Ramachandran 1996). Following this, mirror therapy was identified as a possible treatment modality for unilateral limb impairment (Altschuler *et al*. 1999; Ramachandran and Rogers-Ramachandran 2000). Altschuler *et al*. (1999) demonstrated the potential of mirror therapy in nine chronic stroke survivors, with the quality of motor function improved in those who received mirror therapy. Positive feedback was provided from participants regarding treatment acceptability.

While the mechanisms of neuroplasticity involved are unknown, it is believed that observing movement in the mirror stimulates neurons involved with motor learning and observation, thereby contributing to motor control (Deconinck *et al.* 2015). Other theories link the activation of dormant neurons located in the hemisphere ipsilateral to the paretic limb to improve motor function (Deconinck *et al.* 2015).

A Cochrane review regarding the effectiveness of mirror therapy for the treatment of motor function following stroke was completed in 2012 and included 14 studies; 12 randomised controlled trials (RCT) and two crossover studies (Thieme *et al*. 2012). Thieme *et al*. (2012) reported that mirror therapy may improve motor function and the ability to complete day to day activities. Despite signalling the potential benefit of mirror therapy, there were limitations. The studies included in the review had low sample numbers, wide heterogeneity across participants in terms of time since stroke onset, levels of stroke severity, and variability in mirror therapy programmes.

A more recent Cochrane review published during completion of this PhD, included an additional 49 studies (Thieme *et al*. 2018), highlighting an increase in mirror therapy studies over more recent years. Similar findings to the previous review were found with evidence supporting the use of mirror therapy to improve motor function and performance in ADLs. However, few trials examined the effectiveness of mirror therapy in the early, sub-acute stage of stroke, specifically in the UK, and further research continues to be recommended to build the evidence base.

1.4 Occupational Therapy

Occupational therapists are an integral part of the multidisciplinary stroke rehabilitation team (NICE 2013; Intercollegiate Stroke Working Party 2016), and in the United States stroke survivors account for the largest diagnostic group treated by occupational therapists (National Board for Certified Occupational Therapists 2018). Occupational therapy is concerned with occupational performance and supporting individuals to be as independent as possible in their activities of choice (College of Occupational Therapists 2015). Within stroke rehabilitation the main focus of occupational therapy is improvement in performance of ADLs (Legg *et al*. 2017). Therefore, upper limb treatment often forms the primary focus of occupational therapy due to the implications upper limb impairment poses for engagement in daily activities.

Rehabilitation is generally completed through either restorative or compensatory approaches. Restorative approaches are aimed at driving neuroplastic changes to enhance recovery and activity performance. Compensatory approaches involve changing the activity, environment or how an individual completes a task to facilitate participation (Ivey and Mew 2010). Although the use of activity is central to

occupational therapy treatment, a range of approaches are often used to meet the individual needs of stroke survivors, particularly when a paretic limb is unable to engage functionally in completion of ADLs (Nilsen *et al*. 2010). Therefore, treatments such as mirror therapy can be used by occupational therapists as an adjunct to standard treatment to facilitate neuroplastic changes, for improvement in motor control and subsequent activity performance.

1.5 Pilot studies

Recommended by the UK Medical Research Council guidelines for intervention development, pilot studies form an integral process in the development of large scale RCTs to ensure successful implementation and optimal study validity (Craig *et al*. 2008). Viewed as smaller versions of large-scale studies (Eldridge *et al*. 2016a), pilot studies provide the opportunity to assess recruitment and retention rates as well as staff, time and budget requirements, and to determine intervention delivery, dosage and measurement (Van Teijlingen *et al*. 2001).

The terms ‘pilot’ and ‘feasibility’ are often used interchangeably, with a lack of distinction between the two, leading to inadequate reporting across studies thus limiting their utility (Arain *et al*. 2010). A Delphi study was completed by the Consolidated Standards of Reporting Trials Group to develop a framework for defining pilot and feasibility studies (Eldridge *et al*. 2016a), which informed development of guidelines to improve the reporting of pilot and feasibility studies (Eldridge *et al*. 2016b). Feasibility has been reported as an overarching concept, with randomised and non-randomised pilot studies forming a sub-category, alongside other types of feasibility studies Eldridge *et al*. (2016a).

Additional aspects of feasibility which can be investigated, include whether outcome measures are appropriate and whether the treatment under investigation is acceptable (Lancaster *et al*. 2004). The acceptability of an intervention for stroke survivors is a vital aspect to consider especially with regards to being able to successfully implement a new treatment to ensure adherence and optimal delivery. This is of particular importance in regard to mirror therapy which uses visual feedback to promote recovery as opposed to direct, physical movement of the affected limb.

1.5.1 Outcome assessment

Essential to successful trial implementation is choice of outcome measure. Standardised outcome measures form the basis for evidence-based practice, providing an indicator of patient progress, intervention efficacy and quality of health care (Sullivan *et al*. 2013). Heneghan *et al*. (2017) identified inappropriate outcomes as contributing to the gap in trial outcomes leading to direct patient benefits, with choice of outcome cited as one of the top three methodological research priorities for improving clinical trials (Smith *et al*. 2014). The broad definition of the ICF means its framework can be applied in the assessment of health (WHO 2001). Outcome measures can be categorised across the domains of impairment, activity and participation to enable assessment of complex and subjective aspects of health. Multiple outcome measures are often used in upper limb effectiveness trials to capture the domains of the ICF and demonstrate the multifaceted impact of upper limb impairment following stroke (Santisteban *et al*. 2016).

Outcome measures must demonstrate the psychometric properties of reliability, validity and responsiveness to ensure that they can accurately demonstrate the effects of interventions under investigation (Baker *et al*. 2011). Validity refers to the degree

that an instrument assesses the construct it intends to measure (Mokkink *et al*. 2010). Content validity is the extent the instrument covers all aspects of the construct, as demonstrated by test items. Construct validity is the degree to which an outcome is related to a similar measure, and criterion validity is the degree with which scores correlate with a gold standard outcome measure.

Reliability and agreement assess different aspects of measurement error, evaluating test consistency, and contributing to test validity (Kottner *et al*. 2011; Streiner *et al*. 2015). Reliability is the degree to which an outcome measure differentiates between participants, while agreement demonstrates how identical scores are across repeated measurements. Reliability and agreement can be considered across raters, referred to as inter-rater, and considered for scoring made by one rater at two different time points, referred to as intra-rater. Internal consistency indicates the degree test items correlate with each other and measure the same underlying construct. Finally, where an outcome measure is used at multiple time points to capture change, it must demonstrate responsiveness in order to be sensitive to the changes stroke patients experience throughout rehabilitation (Mokkink *et al*. 2010). Reliability and responsiveness are the two psychometric properties which will be explored further as part of this PhD thesis.

The number of outcome measures available has increased substantially, which can lead to difficulties in choosing an outcome measure (Harrison *et al*. 2013). Issues in compiling systematic reviews and meta-analyses cite heterogeneity in outcomes, impacting on study synthesis and clarity of recommended treatments (Williamson *et al*. 2012). Furthermore, the utility of outcome measures can vary according to time post-stroke and in response to the degree of impairment under investigation. There

are different upper limb assessment tools which capture the complexity of motor control and function according to the impairment and activity domains of the ICF (WHO 2001). With wide heterogeneity, there is no consensus over which outcome measures are suitable for use in the sub-acute stage of stroke (Santisteban *et al*. 2016; Sivan *et al*. 2011), and this area needs further investigation.

1.6 Summary

This chapter has outlined the human and societal burden stroke poses, with stroke identified as a major global healthcare problem. Heterogenous recovery trajectories are found across stroke survivors, with rehabilitation playing a valuable role in enhancing outcomes (Winstein *et al*. 2016). Many individuals fail to re-gain full upper limb functional recovery (Kwakkel *et al*. 2003) and further research is needed to improve current interventions (Pollock *et al*. 2014). Mirror therapy has the potential to drive neuroplastic changes to improve motor recovery and function, and pilot studies are essential for the development of efficacious large-scale trials in this area of rehabilitation. Psychometric evaluation is key to ensuring that outcome measures are fit for purpose and able to accurately demonstrate the effectiveness of interventions. In addition, exploring stroke survivors' viewpoints regarding an intervention has important implications for future intervention and trial design.

1.7 Thesis background

This PhD thesis was completed alongside a pilot RCT designed to investigate the effectiveness of mirror therapy during the sub-acute stage of stroke, defined as within three months of stroke onset (ClinicalTrials.gov: NCT02276729). The pilot study was funded by a grant from the United Kingdom Occupational Therapy Research Foundation in 2014 and led by Dr Alison Porter-Armstrong (APA). Subsequently, a

scholarship was awarded by the Department for the Economy to support the current PhD study completed by Beverley Turtle (BT), which aimed to build the evidence base for upper limb treatment following stroke and explore aspects of feasibility to inform the development of a large-scale multi-site RCT.

1.8 Pilot randomised controlled trial of mirror therapy

The objectives of the pilot RCT were:

1. To establish if it was feasible to recruit participants engaged with in-patient rehabilitation in a sub-acute setting.
2. To evaluate the feasibility of conducting mirror therapy with in-patients, as part of occupational therapy rehabilitation.
3. To evaluate the sensitivity of the outcome measures for use in a fully powered trial and conduct a power calculation.
4. To conduct a preliminary analysis of the data to identify potential treatment gains within and between the 2 groups.
5. To pilot the collection of data to enable cost-consequence analysis to be undertaken as an output of the main RCT.

Participants were recruited across three inpatient rehabilitation sites in the Northern Health and Social Care Trust (NHSCT). The study originally planned to recruit fifty participants; however, this was later adjusted to 40 participants, due to recruitment difficulty. The inclusion criteria for participant selection were**:** 18 years and over; newly admitted inpatient of the rehabilitation ward; diagnosis of CVA in the last three months resulting in upper limb motor loss; able to follow two part spoken or written commands in the English language; upper limb therapy designated as a main portion of goal directed treatment programme; consent to take part in the study.

The exclusion criteria were: patients who had had a previous stroke; gross cognitive impairment or were unable to understand two part spoken/ written commands in English will be excluded.

Once eligible participants consented to take part in the study, individuals underwent block randomisation. Those randomised to the intervention group received two 20-minute sessions of mirror therapy five days per week under the direction of occupational therapy staff and received standard occupational therapy treatment. Participants randomised to the control group received standard occupational therapy treatment.

Participants were assessed using the Functional Independence Measure (FIM™), graded Wolf Motor Function Test (gWMFT), EQ-5D-5L and the Canadian Occupational Performance Measure (COPM) by a research occupational therapist, blinded to treatment allocation. Participants were assessed at baseline following randomisation, every two weeks until discharge, and at three- and six-month follow-up. At discharge, there were 18 participants in the intervention group and 17 participants in the control group, at three months there were 18 participants in the intervention group and 14 participants in the control group. Finally, at six months there were 15 participants in the intervention group and 10 participants in the control group.

1.9 Thesis aim and objectives

The aim of this thesis was to assess the suitability of outcome measures used in a pilot RCT and to explore the acceptability of mirror therapy in sub-acute stroke survivors. The following objectives were identified to meet this aim:

1. To systematically review the literature regarding the application and psychometric assessment of the gWMFT, an upper limb assessment.

2. To investigate the reliability and agreement of the gWMFT within three months of stroke onset.

3. To investigate the responsiveness of the activity-level outcome measures used in the pilot RCT; the gWMFT and FIM.

4. To investigate the responsiveness of the patient-reported outcome measures (PROMs) used in the pilot RCT; the EQ-5D-5L and COPM.

5. To explore the acceptability of mirror therapy with stroke survivors.

Chapter 2

Chapter 2 - A literature review of the application and evaluation of the graded Wolf Motor Function Test

2.1 Abstract

Introduction

Adapted from the Wolf Motor Function Test, the graded Wolf Motor Function Test (gWMFT) is an upper limb activity assessment for use following stroke and brain injury. The aim of this systematic review was to identify and appraise evidence where the gWMFT has been used or has undergone psychometric evaluation.

Method

A systematic review of five databases was conducted to identify studies reporting the gWMFT using a keyword search. Intervention and clinical measurement studies were eligible for inclusion. Data quality was assessed using adapted Critical Appraisal Skills Programme questions and the COnsensus-based Standards for the selection of health Measurement Instruments risk of bias checklist.

Results

Twelve studies, of mostly low quality, were included. Studies included one randomised controlled trial, ten pre-and post-studies and one clinical measurement study. All studies involved participants following stroke. Reliability was the only measurement property assessed in two studies, which were of a ‘doubtful’ and ‘poor’ quality.

Conclusion

Low quality studies impede the ability of clinicians and researchers to best determine the applicability of the gWMFT to patient groups and research contexts. Further exploration of the psychometric properties of the gWMFT is recommended across stroke populations using rigorous design methods.

2.2 Chapter overview

This review was completed to examine articles which described the use and/or psychometric properties of the gWMFT as a measure of upper limb function with individuals presenting with hemiplegia. The researchers involved in the completion of this review are available in Appendix 1. This study was published in the British Journal of Occupational Therapy (Turtle *et al*. 2019) (Appendix 2).

2.3 Introduction

Measurement of upper limb function is complex; the full potential of upper limb function cannot be captured by the use of one outcome measure alone (Ashford *et al*. 2008). Consequently, there is no consensus regarding which outcome measure to use, with choice depending upon intervention (Sivan *et al*. 2011) and sample group (Murphy *et al*. 2015). Upper limb dysfunction can involve paresis, abnormal movement patterns and somatosensory deficits occurring in varying degrees and patterns across stroke survivors (Lang *et al*. 2013). Therefore, outcome measures must, at least in part, reflect the relevant individualised deficits, as well as demonstrate the expected changes resulting from upper limb treatment (Coster 2013; Lang *et al*. 2013). As a result, use of multiple upper limb outcome measures has been found across studies to reflect the myriad outcomes associated with a treatment and the domains of upper limb function impacted (Lang *et al*. 2013; Santisteban *et al*. 2016).

The use of standardised outcome measures is encouraged as part of occupational therapy assessment, with emphasis on those that are reflective of an individual's performance in everyday activities (College of Occupational Therapists 2017). The Wolf Motor Function Test (WMFT) is one of the most frequently reported outcome

measures used in effectiveness trials of upper limb interventions (Santisteban *et al*. 2016). Using the ICF as a framework to categorise outcome measures (WHO 2001), the WMFT is generally reflective of an activity-level outcome measure (Lang *et al*. 2009; Sivan *et al*. 2011; Bushnell *et al*. 2015; Santisteban *et al*. 2016). Activity-level outcome measures are often viewed as integral to demonstrating meaningful patient outcomes, offering direct insight into their ability to complete everyday tasks (Lang *et al*. 2013; Pike *et al*. 2018).

The WMFT was designed to assess the upper limb motor function of individuals with hemiplegia following stroke or brain injury (Taub *et al*. 2011). Originally called the Emory Motor Test (Wolf *et al*. 1989), the assessment consisted of 21 items. Since then, this assessment has undergone further adaptation and now consists of 15 test items for which participants are scored on their speed and quality of performance, and two strength tasks (Taub *et al*. 2011). Test items assess joint-specific movements, progressing through to more difficult items requiring integration for completing functional tasks. The inclusion of functional tasks has led test authors to surmise this laboratory-based assessment may mirror an individual's functional use of their more affected upper limb in everyday life (Morris *et al*. 2001).

The psychometric properties of an outcome measure are integral for determining their suitability for use with a patient group. Outcome measures must be reliable and valid to accurately reflect patient progress and demonstrate treatment effectiveness (Sullivan *et al*. 2013). The WMFT has undergone extensive evaluation of its psychometric properties to support its use in stroke rehabilitation trials. The WMFT has demonstrated high levels of inter-rater reliability (intraclass correlation coefficient (ICC) >0.9) and test-retest reliability (r >0.9) for functional ability and performance time

when used with individuals who are more than 12 months' post-stroke (Morris *et al*. 2001). Adequate responsiveness was found for the WMFT when used with individuals at least six months' post-stroke, with higher levels of responsiveness found for functional ability scores (Hsieh *et al*. 2009). Additional studies have quantified construct validity, criterion validity and predictive validity for the WMFT (Wolf *et al*. 2001; Hsieh *et al*. 2009). The WMFT has been validated for use with a chronic stroke population and a multidisciplinary panel recommended the WMFT as a secondary outcome measure for use in intervention studies (Bushnell *et al*. 2015). In a systematic review of upper limb outcome measures used in stroke research, the WMFT was the second most reported outcome measure across 477 studies and the most commonly reported activity-level outcome measure (Santisteban *et al*. 2016).

Although a popular tool, the WMFT has limitations. It is an outcome measure designed to capture the upper limb capabilities of those with mild to moderate deficits. Thompson-Butel *et al.* (2015) found floor effects when the WMFT was used with stroke survivors with greater levels of upper limb dysfunction. In this study just over 30% of participants with severe upper limb impairment were unable to complete any test item and achieved the maximum performance time score of 120 seconds (Thompson-Butel *et al*. 2015). Lin *et al*. (2009a) similarly found floor effects for functional ability scores when used in the earlier stages of stroke; 17% of participants achieved a score of zero on the functional ability scale (FAS) at 14 days post stroke. Therefore, to capture the upper limb function of individuals with a greater degree of impairment the test authors also developed the graded Wolf Motor Function Test (gWMFT) (Constraint Induced Research Therapy Group 2002).

Description of the gWMFT

The gWMFT consists of 13 items, and progresses hierarchically, in a similar pattern as the WMFT (Constraint-Induced Research Therapy Group 2002). However, each item consists of two levels (level A and level B), meaning the item can be adjusted to an individual's level of ability accordingly (Appendix 3). A manual is available for the gWMFT with detailed instructions on test administration and scoring, promoting standardisation (Constraint Induced Movement Therapy Research group 2002).

Scoring the gWMFT

Individuals are scored based on their quality of movement and speed of performance. Their quality of movement is scored using the FAS. The FAS for the gWMFT is an eight-point ordinal scale, with scores ranging from zero, representing no active movement, through to seven, representing normal movement (Appendix 4). Test authors recommend videotaping the assessment and scoring functional ability at a later time to improve accuracy (Constraint Induced Therapy Research Group 2002).

Scores for performance time and functional ability are determined by the level of item completed. For items completed at Level A individuals must complete the item within 30 seconds and can score between four and seven on the FAS. If an Individual is unable to complete a test item within 30 seconds, they are then able to attempt the level B version for that item. Individuals receive an extra 60 seconds added onto their performance time score for items completed at level B. In addition, individuals are only able to score between zero and three on the FAS with a maximum time of 120 seconds. The test authors recommend the median time and the mean functional ability scores are reported as summary scores for each patient.

However, in comparison to the WMFT, there remains limited uptake of the gWMFT in clinical trials and limited evaluation of its psychometric properties. Such studies are required to determine the suitability of the outcome measure across patient groups and upper limb interventions. Outcome measurement is an important attribute to consider as part of evidence-based practice. Therefore, a systematic literature review encompassing examination of the gWMFT is warranted to build on the evidence base for upper limb assessment following stroke.

2.4 Aim and objectives

2.4.1 Aim

This chapter aimed to explore how the gWMFT has been utilised and reported in the literature.

2.4.2 Objectives

1. To identify and evaluate studies where the gWMFT has been used as a primary and/or secondary outcome measure.
2. To summarise how the gWMFT has been reported.
3. To identify and evaluate evidence for the measurement properties of the gWMFT.

2.5 Method

A systematic literature review was completed by searching the following electronic databases: CINAHL (Cumulative Index to Nursing and Allied Health Literature) (1937-October 2018), Ovid MEDLINE (1949-October 2018), AMED (1985- October 2018), PsycINFO (1806- October 2018) and Pubmed (1947- October 2018). A search

strategy was formulated using the following keywords in combination: “graded wolf motor function test” OR gwmft (Appendix 5). This systematic literature review was completed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher *et al*., 2009). This approach was used to enable a transparent and structured search of the literature.

2.5.1 Inclusion and exclusion criteria

Articles were included if they were in the English language and featured the gWMFT as a primary or secondary outcome measure. Due to the limited review scope, and the focus on the use and application of the gWMFT, any studies regardless of patient population, clinical intervention or methodological design were included. Review articles and those where only an abstract was available were excluded.

2.5.2 Study selection

Search results were transferred to the Refworks reference management programme and duplicates were removed. Titles and abstracts for all retrieved studies were screened and examined for any reference to the gWMFT by BT. The reference list for each relevant publication was also searched, which led to the retrieval of one additional article. The full-text format of papers were reviewed where the outcome measures were not reported in the abstract, and for further examination of study criteria by BT. The studies retrieved were independently checked by APA and Dr May Stinson (MS), and eligibility confirmed. Differences in opinion were resolved through discussion between BT, APA and MS. One of the full text papers retrieved involved the grade 5 WMFT (Bowman *et al*., 2006) and the consensus decision was made to exclude this article.

2.5.3 Data collection and analysis

Data for all included studies were extracted by BT and recorded using Excel spreadsheets. The data extracted included participant characteristics (age, gender); time post-stroke; study design; intervention applied; and psychometric properties. Aspects of the gWMFT extracted included version reported; and scoring attributes.

The quality of included studies was assessed using an adapted version of the Critical Appraisal Skills Programme (CASP) for cohort studies (CASP 2018), which included questions used to assess the quality of studies examining outcome measures (Jerosch-Herold, 2005) (Table 2.1). This combination was chosen due to the varied type of studies found and the focus of the review on outcome measurement.

Table 2.1 Adapted CASP questions

1.	Did the study address a clearly focused issue?
2.	Was the sample recruited in an acceptable way? Is the sample size adequate (is there a power calculation)?
3.	Is the instrument described and accurately measured to minimise bias?
4.	Are the testers trained in test administration?
5.	Have the authors identified and taken into account confounding factors?
6.	Did the study address a clearly focused issue?

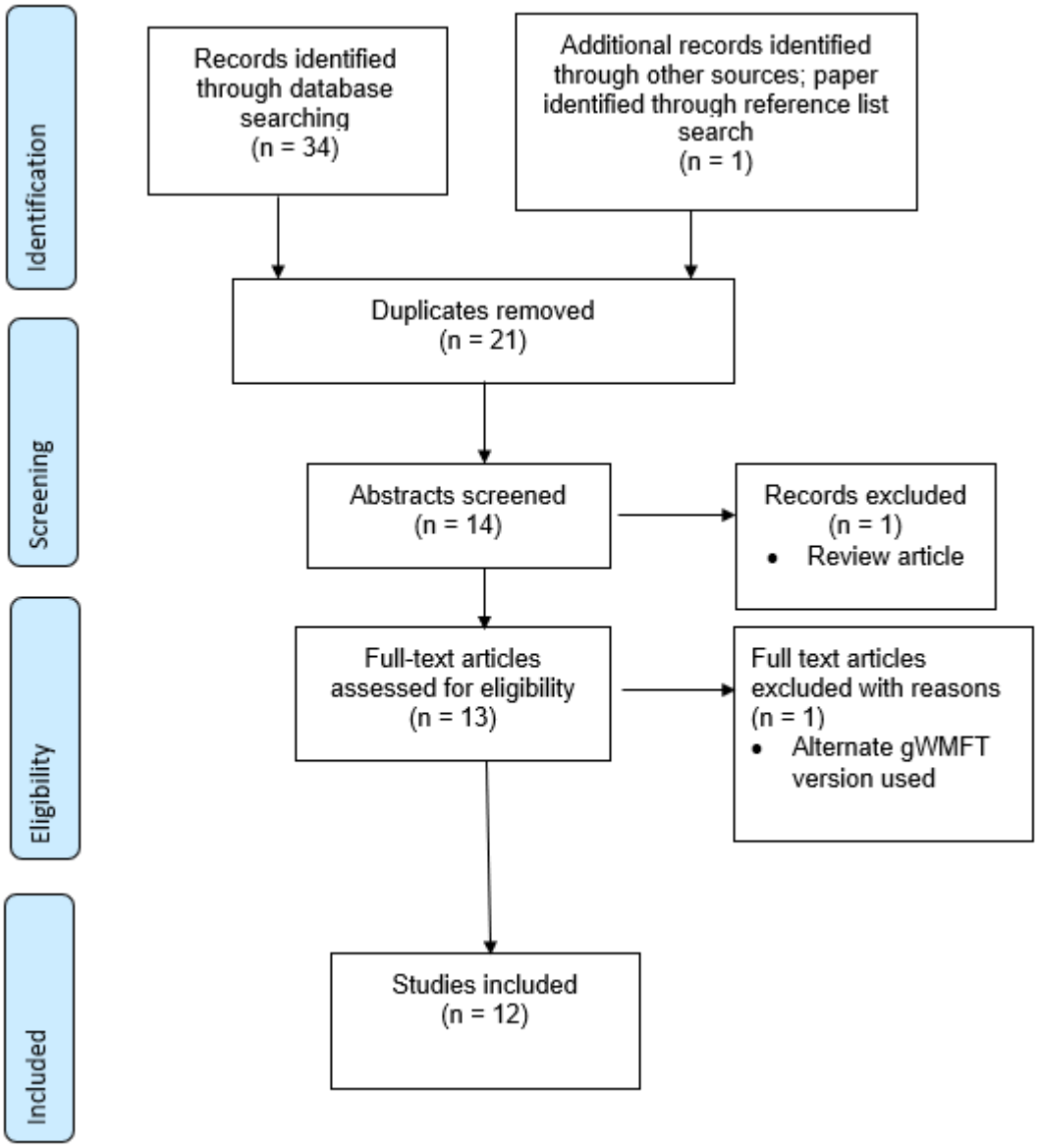
Where studies examined the psychometric properties of the gWMFT, the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist was used (Mokkink *et al*., 2018). Designed originally for health-related patient-reported outcomes, the COSMIN can be applied to observer-reported outcome measures, to aid selection and reporting (Hernaez 2015). Using the risk of bias checklist, the measurement property was evaluated and rated ‘good’, ‘fair’, ‘doubtful’, ‘poor’ or ‘not applicable’. The lowest rating achieved determined the overall methodological quality for each applicable study. Each study was assessed on these attributes by BT, and scores were agreed upon with APA and MS.

2.6 Results

Thirty-four articles were identified from the search of electronic databases, with one article identified from the reference section of other studies and are summarised in the PRISMA flow diagram in Figure 2.1. Following the removal of duplicates (n=21), the titles and abstracts of the 14 remaining articles were screened, which led to the removal of a review paper (n=1). The full-text articles for 13 studies were retrieved, and one study was removed because it did not meet the inclusion criteria. The

removed study involved the use of a different version of the WMFT, named the Grade 5 WMFT (Bowman *et al*. 2006). There were 12 studies included in this review.

Figure 2.1 Summary of the literature review search using the PRISMA group flow chart (Moher *et al*., 2009)



2.6.1 Characteristics of included studies

The characteristics of the included papers are summarised in Table 2.2. Most studies were performed in the United States of America (n=9), two were completed in India, and one was completed in Brazil. Eleven of the studies were intervention based, and these were mostly of a pre- and post-test design. All interventions were aimed at improving upper limb function in individuals following stroke. One study was an inter-rater reliability and agreement study of the Brazilian Portuguese version of the gWMFT (Pereira *et al*. 2015).

The sample sizes for included studies were low; nine studies consisted of 20 participants or less (Bonifer and Anderson 2003; Bonifer *et al*. 2005; Flinn *et al*. 2009; Iwamuro *et al*. 2011; Triandafilou *et al*. 2011; Demirtas-Tatlidede *et al*. 2015; Triandafilou *et al*. 2014b; Pereira *et al*. 2015; Fischer *et al*. 2016). The study with the largest sample size was completed by Arya *et al*. (2012) which consisted of 103 participants. While the study completed by Triandafilou and Kamper (2014a) consisted of 27 participants those participants were divided further into two groups according to time post-stroke and analysed separately. Twelve of the participants were two to six months' post-stroke and 15 participants were more than six months' post-stroke (Triandafilou and Kamper 2014a).

All studies were completed with individuals following stroke, and more than half of the studies were completed with individuals six months or more post-stroke (Bonifer and Anderson 2003; Bonifer *et al*. 2005; Flinn *et al*. 2009; Iwamuro *et al*. 2011; Triandafilou *et al*. 2011; Demirtas-Tatlidede *et al*. 2015; Pereira *et al*. 2015). Five of the studies came from the same research team, using a device called the X-Glove to passively

stretch the fingers of the more affected hand (Iwamuro *et al*. 2011; Triandafilou *et al*. 2011; Triandafiou and Kamper 2014a; Triandafilou *et al*. 2014b; Fischer *et al*. 2016).

Table 2.2 Characteristics of the studies included in the review

Author (year)	**Patient (n)**	**Country**	**Time post stroke**	**Age, years (mean ±SD), Gender**	**Type of study**	**Intervention**	**Additional upper limb assessments included**
Anandabai and Gupta (2012)	30	India	3-4 months	Not reported 26 male, 6 female	Pre- and post-study	Bimanual (n=15) and unimanual (n=15) functional practice	Fugl-Meyer Assessment
Arya *et al*. (2012)	Intervention: 51 Control: 52	India	4-24 weeks	Intervention: 51.67 ±7.96 29 male, 22 female Control: 50.21 ±7.60 33 male, 19 female	RCT	Meaningful task-specific training	Action Research Arm Test Fugl-Meyer Assessment
Bonifer and Anderson (2003)	1	USA	15 years	53 1 female	Case report	Constraint-induced movement therapy	Fugl-Meyer Assessment Motor Activity Log
Bonifer *et al*. (2005)	20	USA	>12 months	57.5 ±16.6 13 male, 7 female	Pre- and post-study/Psychometric study	Constraint-induced movement therapy	Fugl-Meyer Assessment Motor Activity Log
Demirtas-Tatlidede *et al*., (2015)	10	USA	>1 year	59.5 ±11 4 male, 6 female	Pre- and post-study.	Contra-lesional repetitive transcranial magnetic stimulation	Fugl-Meyer Assessment Hand strength tests Modified Ashworth Scale
Fischer *et al*. (2016)	15	USA	2-6 months	63 ±12 10 male, 3 female	Pre- and post-study	Passive cyclical finger stretching with active-assisted task-oriented	Action Research Arm Test

						training using an orthotic glove	Chedoke Arm and Hand Inventory Fugl-Meyer Assessment Hand strength Motor Activity Log
Flinn *et al*. (2009)	1	USA	15 months	48 1 female	Case report	Robot-assisted therapy	Active range of motion Fugl-Meyer assessment Motor Activity Log
Iwamuro *et al*. (2011)	5	USA	≥9 months	54 ±11 4 male, 1 female	Pilot pre- and post-study	Passive finger extension using an orthotic glove	Active range of motion Box and Block Test
Pereira *et al*. (2015)	10	Brazil	>6 months	53.2 ±11.39 6 male, 4 female	Psychometric study	N/A	N/A
Triandafilou *et al*. (2011)	15	USA	≥6 months	57 ±8 7 male, 8 female	Pre- and post-study	Prolonged and repetitive passive finger stretching using an orthotic glove	Hand strength tests Grip termination time
Triandafilou and Kamper (2014a)	Sub-acute: 12 Chronic: 15	USA	Subacute: 2-6 months Chronic: >7months	Sub-acute: 53 ±6 6 male, 6 female Chronic: 57 ±8 7 male, 8 female	Pre- and post-study	Passive cyclical finger stretching using an orthotic glove	Box and Block Test Hand strength tests
Triandafilou *et al*. (2014b)	13	USA	2-6 months	51 ±12 6 male, 7 female	Pre- and post-study	Static finger stretching/passive finger stretching/rest using an orthotic glove	Hand strength tests Grip termination time

Abbreviations: SD, standard deviation; RCT, randomised controlled trial.

2.6.2 Quality of included studies

The scoring of studies, which used the adapted version of the CASP tool (CASP 2018) and questions used for the evaluation of outcome measurement studies (Jerosch-Herold 2005), found that they were generally of low quality (Table 2.3). One of the studies was a RCT. It was the only study to complete a sample size calculation, an intention to treat analysis and an examination of between-group differences in baseline characteristics (which included age, gender and time since stroke) (Arya *et al*. 2015).

Most pre- and post-test study designs did not account for how participants were recruited and did not report how bias was reduced. Studies with low sample sizes, and potentially insufficient statistical power were used to determine treatment effects. Most studies did not describe the gWMFT adequately and did not report how the gWMFT was administered and scored. The third most common reason for allocating low scores was due to lack of clarity regarding how test authors reduced confounding factors, including whether participants and assessors were blinded.

A key example of this was in the study by Anandabai and Gupta (2012). This study reported participants who were randomised into one of two groups in the abstract. However, in the method section for this study it was detailed that participants were sampled by convenience into one of two groups, rendering the study liable to allocation bias. There was no detail provided regarding the administration and scoring of the outcome measures delivered. Also, participant demographic information was not reported for the two groups and not controlled for in the subsequent analysis.

There were two case reports which clearly described the participant's health status and the intervention delivered (Bonifer and Anderson 2003; Flinn *et al*. 2009).

Table 2.3 Critical appraisal scoring for each reviewed article

Author (year)	**Address focused issue**	**Acceptable recruitment**	**Adequate sample size**	**Instrument described and accurately measured**	**Testers trained in administration**	**Confounding factors identified and taken into account**
Anandabai and Gupta (2012)	Y	CT	N	N	CT	N
Arya *et al*. (2012)	Y	Y	Y	CT	Y	Y
Bonifer and Anderson (2003)	Y	N/A	N/A	Y	Y	N/A
Bonifer *et al*. (2005)	Y	Y	N	Y	Y	Y
Demirtas-Tatlidede *et al*. (2015)	Y	CT	N	N	CT	N
Fischer *et al*. (2016)	Y	CT	N	N	CT	CT
Flinn *et al*. (2009)	Y	N/A	N/A	N	Y	N/A
Iwamuro *et al*. (2011)	Y	CT	N	N	N	CT
Pereira *et al*. (2015)	Y	CT	N	Y	CT	CT
Triandafilou *et al*. (2011)	Y	CT	N	N	CT	CT
Triandafilou and Kamper (2014a)	Y	CT	N	N	CT	CT
Triandafilou *et al*. (2014b)	Y	CT	N	N	CT	CT

Abbreviations: Y, yes; N, no; CT, cannot tell; N/A, not applicable.

2.6.3 Quality of studies reporting psychometric evaluation

Two studies investigated aspects of the psychometric properties of the gWMFT (Bonifer *et al*. 2005; Pereira *et al*. 2015), and methodological quality was assessed using the COSMIN risk of bias checklist (Table 2.4). Although not an aim of the study by Bonifer *et al*. (2005), the intra-rater reliability of the 14-item gWMFT was completed as part of their intervention study. Intra-rater reliability for scoring functional ability was reported (Table 2.4), indicating a good level of reliability. Paired raters completed scoring of functional ability; a physiotherapist and an occupational therapist viewed recorded videos of participants completing the gWMFT. A high level of detail was reported regarding the training of raters. However, the time interval between reviewing recorded videos was not reported and as such, it is not clear how bias was minimised. In addition, it is not stated how a final score was achieved between raters. The statistical analysis used was a Pearson product moment correlation, which is not an advised method of analysis according to the COSMIN (Mokkink *et al*. 2018). Using the COSMIN risk of bias checklist this study received a rating of 'poor' due to the limited methodological detail reported regarding the time interval between scoring sessions and how a final score was achieved between raters (Table 2.4).

The study by Pereira *et al*. (2015) was an inter-rater reliability study of the Brazilian Portuguese version of the 13-item gWMFT. Inter-rater reliability and agreement for both functional ability and performance time were reported (Table 2.4). An excellent level of inter-rater reliability was found for scoring functional ability (intraclass correlation coefficient (ICC)=0.98 [95% confidence interval [CI]=0.92-0.99]) and performance time (ICC=0.99 [95% CI=0.95-1.00]). An adequate amount of agreement was found for scoring functional ability (limits of agreement were between -0.68 and

0.6). Although not noted by test authors, the limits of agreement for scoring performance time indicated an inadequate level of agreement, with limits between -0.68 and 16.1 seconds. The mean difference between raters was 5.5 seconds. Two raters independently administered the gWMFT and scored performance time and functional ability through direct observation to 10 participants more than six months post-stroke. There was an adequate amount of time between scoring by the individual raters, approximately two weeks. However, the type of ICC performed was not reported. In line with guidelines for scoring the COSMIN risk of bias checklist (Mokkink *et al*. 2018), this study received a rating of 'doubtful,' due to lack of clarity in reporting aspects of the reliability analysis (Table 2.4).

Table 2.4 COnsensus-based Standards for the selection of health Measurement INstruments risk of bias checklist to assess the methodological quality of the included reliability studies

Design requirements	**Bonifer *et al*. (2005)**	**Pereira *et al*. (2015)**
Were patients stable in-between measurements?	F	F
Time interval appropriate?	D	G
Conditions similar for both measurements?	F	F
Reliability analysis		
ICC for continuous scores?	F	F
Kappa for dichotomous, ordinal, or nominal scores?	N/A	N/A
Weighted kappa for ordinal scores?	N/A	N/A
Weighting scheme described for ordinal scores?	N/A	N/A
Any important flaws in design?	P	D
Agreement analysis		
Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated for continuous scores	N/A	G
Percentage (positive and negative) agreement calculated for nominal/ordinal scores	N/A	N/A
Any important flaws in design	N/A	D
Final score	P	D

Abbreviations: G, good; F, fair; D, doubtful; P, poor; N/A, not applicable

2.6.4 Reporting of the gWMFT

The version of the gWMFT used by the included studies is demonstrated in Table 2.5.

Although not reported, Pereira *et al*. (2015) adapted 13 items from the gWMFT into

Brazilian Portuguese, indicating the use of the latest version. Similarly, Iwamuro *et al*. (2011) did not reference the gWMFT manual in their study. However authors report the gWMFT consisted of 13 items, indicating the use of the 2002 version. The 2000 version of the gWMFT referenced by Bonifer and Anderson (2003) and Bonifer *et al*. (2005) had an additional item, 'drop golf ball or washcloth' and required participants to stand to complete items 10 to 14. Fischer *et al*. (2016) cite the study by Bonifer and Anderson (2003) indicating the use of the 2000 version.

In the study by Flinn *et al*. (2009), it was not clear if a variation of the gWMFT was used. The study authors reported that tasks requiring fine motor control were removed, the number of tasks requiring gross motor function were reduced and tasks which required pronation and supination were increased. It was also not clear which version was used due to insufficient reporting in the studies by Anandabai and Gupta (2012) and Demirtas-Tatlidede *et al*. (2015).

Table 2.5 Reporting of the gWMFT

Author (year)		**Domains of gWMFT scored**	

	13-item/ 14-item gWMFT	FAS	Performance time	Description of scoring criteria
Anandabai and Gupta (2012)	Not reported	✓	✓	Not reported
Arya *et al*. (2012)	13-item	✓	✓	✓
Bonifer and Anderson (2003)	14-item	✓	✓	✓
Bonifer *et al*. (2005)	14-item	✓	✓	Not reported
Demirtas-Tatlidede *et al*. (2015)	Not reported	Not reported	Not reported	Not reported
Fischer *et al*. (2016)	Not clear, cited Bonifer and Anderson (2003)	✓	✓	Not reported
Flinn *et al*. (2009)	Not clear		✓	Not reported
Iwamuro *et al*. (2011)	Not clear, reported 13 items		✓	Not reported
Pereira *et al*. (2015)	13-item	✓	✓	✓
Triandafilou *et al*. (2011)	14-item		✓	✓*
Triandafilou and Kamper (2014a)	14-item		✓	✓*
Triandafilou *et al*. (2014b)	14-item		✓	✓*

Note: ✓, reported; ✓* reported with adaptations

Minimum level of function

Eight of the studies required research participants to have a minimum level of function, as part of their eligibility criteria. A significant level of hand impairment was required by Iwamuro *et al*. (2011), Triandafilou and Kamper (2014a) and Triandafilou *et al*. (2014b) and Fischer *et al*. (2016) which was determined by the Chedoke McMaster Stroke Scale for the hand.

It was unclear what criteria were applied in the study by Anandabai and Gupta (2012) and further exploration was not possible. Arya *et al*. (2012) used the Brunnstrom stages of arm recovery and those at stages two to five were eligible for their study. Stage two of the Brunnstrom stage of recovery indicates the presence of basic limb synergies and minimal movement responses and at stage five the upper limb is capable of more complex movement patterns (Safaz *et al*. 2009). The Fugl-Meyer upper extremity scale was used by Demirtas-Tatlidede *et al*. (2015), participants were required to score ≤16, indicating a severe level of impairment.

Measurement of the gWMFT

There were variations in how studies reported the scoring criteria for the gWMFT (Table 2.5), with many not stating the differences in scoring according to level of item completed for functional ability and/or performance time (Bonifer *et al*. 2005; Flinn *et al*. 2009; Iwamuro *et al*. 2011; Anandabai and Gupta 2012; Demirtas-Tatlidede *et al*. 2015; Fischer *et al*. 2016). Only studies by Arya *et al*. (2012), Bonifer and Anderson (2005) and Pereira *et al*. (2015) detailed how functional ability and performance time were scored according to the level of item completed. Iwamuro *et al*. (2011), Triandafilou *et al*. (2011), Triandafilou and Kamper (2014a), Triandafilou *et al*. (2014b) assessed participants on only three tasks of the gWMFT. The three tasks included: lifting a pen, lifting cotton balls and lifting a washcloth (Iwamuro *et al.* 2011; Triandafilou *et al*. 2011; Triandafilou and Kamper 2014a). Triandafilou *et al*. (2014b) reported the use of three hand-specific items from the gWMFT, which were not detailed.

Study authors additionally adapted scoring criteria. In the study by Triandafilou and Kamper (2014a) participants received an additional 60 seconds for not using the

appropriate grasp, with a maximum of 120 seconds, and the sum score for the three tasks was reported. Triandafiou *et al*. (2014b) reported each task was completed three times for each assessment period, the averages for each task were then summed and used to summarise each assessment session. Triandafilou *et al*. (2011) also reported that each task was completed in sets of three, with a maximum time of 60 seconds allowed. Study authors did not report what summary score was used, which then underwent a logarithmic transformation (Triandafilou *et al*. 2011). In the study by Arya *et al*. (2012) each participant's performance time score was that of their less affected arm subtracted from the score for their more affected arm. Flinn *et al*. (2009) used the summation of performance time scores as a summary score.

Additional upper limb outcome measures included

Most studies included additional upper limb outcome measures; these are demonstrated in Table 2.5. The Fugl-Meyer upper extremity scale was most commonly reported, followed by measures of hand and arm strength.

2.7 Discussion

This systematic literature review has provided an overview of how the gWMFT has been reported in intervention studies and includes the limited assessment of its psychometric properties. Although the gWMFT was designed for individuals following stroke or brain injury all included studies involved stroke survivors only. The assessment of the quality criteria of included studies demonstrated most were of a poor or dubious quality due to the inconsistent administration and scoring of the gWMFT. As a result of minimal investigation, evaluation of the psychometric properties of the gWMFT was not used to guide the review. The current review identified one

inter-rater reliability study of the Brazilian Portuguese version of the gWMFT (Pereira *et al*. 2015) and an intra-rater reliability study of the gWMFT FAS (Bonifer *et al*. 2005).

Hierarchies of evidence are used within healthcare to aid the interpretation of research studies of varying designs (Evans 2003). Within studies which evaluate the effectiveness of interventions, systematic reviews and RCTs are viewed as delivering the highest quality of evidence and interpreted as trusted contributors to evidence-based practice (Evans 2003). However, in determining intervention effectiveness, RCTs must demonstrate precision in their design, conduct, and analysis to meet gold standard criteria (Higgins *et al*. 2011).

There was one RCT included in the review which examined the effectiveness of meaningful task-specific training using the gWMFT as a secondary outcome measure (Arya *et al*. 2012) and scored highly across all quality criteria. Demonstrating the effectiveness of an intervention under investigation relies on the ability of the outcome measure to convey treatment effects accurately, and measurement of appropriate outcomes is a top methodological priority for improving clinical trials (Smith *et al*. 2014). In the study by Arya *et al*. (2012) the gWMFT was explained in detail; the scoring criteria for performance time and descriptors of the ordinal scale used to score functional ability were reported aiding study replication.

The remaining intervention studies included in the review consisted of pre- and post-test designs. Non-randomised controlled trials are viewed as liable to increased bias and as such register lower on the hierarchy of evidence (Higgins *et al*. 2011). The description of outcome measures included was generally of low quality, with some simply stating the gWMFT was used and no further examination provided (Iwamuro *et al*. 2011; Anandabai and Gupta 2012; Fischer *et al*. 2016). Also, how test authors

reduced confounding factors were infrequently reported, with many not reporting whether assessors were blinded.

Case reports are descriptive and provide little insight into the efficacy of a treatment and are rated poorly in the hierarchy of evidence (Evans 2003). However, the case report by Bonifer and Anderson (2003) provided an in-depth description of the training provided to raters and of the gWMFT administration, scoring and test items included. While this study may not score highly in determining effectiveness, this study scored highly in the current review due to how the gWMFT was described and administered. In contrast, the case study by Flinn *et al*. (2009) poorly described the gWMFT and inaccurately reported the study completed by Bonifer *et al*. (2005) as an inter-rater reliability study.

2.7.1 Psychometric properties of the gWMFT

Reliability was the only measurement property assessed for the gWMFT. While the results reported a high level of inter- and intra-rater reliability, the methodological quality of these studies were of a low standard.

In the intra-rater reliability study by Bonifer *et al*. (2005), two raters scored functional ability using the videotapes of participants completing the gWMFT. Although the level of training provided was detailed, it was not clear how the two raters came to a final agreed score for each participant. This study scored poorly on the COSMIN risk of bias checklist for not reporting the time interval between the repeated measurements of functional ability. Reporting the time interval is integral to determine that a long enough period has elapsed to prevent the raters from remembering their previous scores (Terwee *et al*. 2007; Kottner *et al*. 2011). Finally, this study reported reliability

using Pearson's product moment correlation. This is not an advised method of analysis due to lack of further exploration of differences in scores between the two-time points (Streiner and Kottner 2014; Mokkink *et al*. 2018).

The study by Pereira *et al*. (2015) investigated the Brazilian Portuguese version of the gWMFT, which limits its applicability to the English language version. This study was scored 'doubtful' on the COSMIN risk of bias due to lack of clarity regarding which type of ICC was completed. How an ICC value is interpreted depends on the type used for the analysis, with assumptions made regarding the number of raters involved and how raters score participants (Kottner and Dassen 2008).

The level of agreement was also assessed by Pereira *et al*. (2015). Bland and Altman plots were used to determine the limits of agreement for scoring functional ability and performance time. Approximately 95% of the difference in scores between the raters will lie within the limits of agreement (Bland and Altman 1999). Ideally, this should be close to zero, indicating minimal differences. A large degree of measurement error was found for scoring performance time, illustrated through wide limits of agreement in the Bland and Altman plot, with a mean difference of 5.5 seconds between raters (Pereira *et al*. 2015).

Although Fritz *et al*. (2009) assessed the minimal detectable change for performance time of the WMFT, Fritz *et al*. (2009) reported that change of at least 0.7 seconds indicated an improvement in performance. The wide degree of measurement error as demonstrated by Pereira *et al*. (2015) for the gWMFT, would make it difficult to discern whether a change in a participant's score was the result of an actual change in recovery or the result of measurement error. Continued validation of the gWMFT is

necessary to determine its ability to measure and document change in upper limb function accurately.

2.7.2 Application of the gWMFT

Across studies, there was heterogeneity in the version of the gWMFT used, how it was applied and scored. Most of the included studies did not provide an adequate description of the complexity involved in delivering and scoring the gWMFT. In comparison to the 12 articles suitable for inclusion in this review, a search of the WMFT using Ovid MEDLINE elicited 384 studies. While increased use of the WMFT is to be expected, there remains a wide disparity in uptake between the two outcome measures. Poor reporting in studies of low quality is likely to play a role in whether clinicians or researchers choose to use the gWMFT.

The gWMFT was used with a variety of upper limb interventions including constraint-induced movement therapy, robotic therapy, repetitive task training, and transcranial magnetic stimulation. In addition to the intervention under investigation, the amount of time which has passed following stroke and level of upper limb impairment can play a factor in the choice of outcome measure (Sivan *et al*. 2011). Most studies assessed stroke survivors more than six months post stroke and required participants to demonstrate severe upper limb impairment, as part of their inclusion criteria.

The gWMFT could provide an appropriate alternative to the WMFT for use in the earlier stages of stroke and with those with a greater degree of impairment. However, included studies did not report floor and ceiling effects which would gauge the sensitivity of the gWMFT to measure severe upper limb deficits accurately. Further

studies assessing the reliability, validity, and responsiveness of the gWMFT are required to ascertain its appropriateness across interventions and level of impairment.

Most studies assessed participants on more than one outcome measure, which has been identified across upper limb intervention trials (Santisteban *et al*. 2016). The Fugl-Meyer upper extremity scale was the most widely reported additional outcome measure cited in just over half of the included studies. The Fugl-Meyer is one of the most widely reported (Murphy *et al*. 2015; Santisteban *et al*. 2016) and recommended outcome measures for use in upper limb research (Bushnell *et al*. 2015). However, the Fugl-Meyer upper extremity scale is an assessment categorised at the body function domain of the ICF (Gladstone *et al*. 2002) and is limited regarding how the scores relate to a stroke survivor's ability to complete everyday activities. The second most commonly reported outcome measure involved measurements of hand strength. Similar to the Fugl-Meyer upper extremity scale, assessment of hand strength is categorised at the level of body function. These provide complimentary and objective outcome measures to consider alongside activity-level outcome measures.

2.7.3 Adaptation of the WMFT

Through the preparation of this review, another adaptation of the WMFT was found called the Grade 5 WMFT (Uswatte *et al*. 2018). This test is comparable to the gWMFT in that each item consists of two levels, with similar scoring criteria. The Grade 5 version consists of 10 items, and the mean log score was reported for performance time. The gWMFT was developed to assess the upper limb motor function of individuals with moderate to severe deficits, while the Grade 5 WMFT was developed to assess the motor function of individuals with severe deficits.

The University of Alabama, Birmingham, United States of America, devised a system for classifying the minimum amount of active range of motion at each joint of the upper limb, which can be used to determine the appropriate WMFT to use (Uswatte and Taub 2013; Uswatte *et al*. 2018). This was established by the same research team responsible for the development of the WMFT. Using this system individuals classified under grade 2/3 with a mild to moderate level of impairment, would be appropriate to complete the original WMFT. Individuals classified on grades 3/4 would be appropriate to complete the gWMFT and individuals classified under grade 5 would be appropriate to complete Grade 5 WMFT (Uswatte and Taub 2013; Uswatte *et al*. 2018).

None of the studies in this review used this classification system as part of their inclusion criteria, with seven studies using three different standardised outcome measures to determine the minimum level of function required. This level of heterogeneity concerning the level of ability for which the gWMFT is an appropriate outcome measure limits its clinical utility.

2.7.4 Limitations

This review was limited by the small number of studies identified, with wide heterogeneity impacting data synthesis, which limited the comparisons that could be made. Quality appraisal of the studies was completed using an amalgamation of two quality appraisal tools (CASP, 2018; Jerosch-Herold 2005). Although not a standardised tool, this was created to cope with the heterogeneity of the included studies.

2.7.5 Implications for future research

While there are studies which have reported the gWMFT to a high standard, these were in the minority. Therefore, future attention should be given to the development of high-quality research measuring upper limb function in stroke survivors using the gWMFT. Future research should also focus on the applicability of the gWMFT to stroke survivors at different stages of their recovery to ascertain its clinical utility.

There was wide heterogeneity in how the gWMFT was reported; some studies reported using the gWMFT where participants were assessed on three of the test items. The study authors did not validate the use of these three items, and to use them when the gWMFT itself has not been sufficiently validated is questionable. To rectify this priority should be given to assessing the measurement properties of the gWMFT. The COSMIN checklist can be used as a guide when investigating the psychometric properties of the gWMFT. Guidelines also exist for reporting reliability and agreement studies to aid study transparency and improve the quality of reporting (Kottner *et al*. 2011).

2.8 Conclusion

This review has demonstrated that while the gWMFT has limited uptake, researchers are continuing to use this outcome measure with limited evaluation of its measurement properties. To date, there has been an inter-rater reliability and agreement study of performance time and functional ability for the Brazilian Portuguese version. A similar evaluation has not been completed for the English language version. While the intra-rater reliability study completed by Bonifer *et al*. (2005) was a step in the right direction, this study lacked rigor and only considered the reliability of functional ability. Improvements in upper limb interventions is a key priority for stroke survivors, carers

and health professionals (Pollock *et al*. 2012). The gWMFT has the potential to extend the applicability of the WMFT and generate meaningful results concerning the recovery of upper limb function in individuals undergoing stroke rehabilitation. Therefore, the following chapters seek to investigate some of the measurement properties of the gWMFT when used with stroke survivors engaging in mirror therapy.

Chapter 3

Chapter 3 – Inter- and intra- reliability and agreement of the graded Wolf Motor Function Test in sub-acute stroke

3.1 Abstract

Introduction: Clinical utility of the graded Wolf Motor Function Test (gWMFT) is limited by a lack of psychometric evaluation and the requirement to video record participant performances for scoring purposes. This study aimed to (1) assess whether video recording is required through examination of inter-rater reliability and agreement; and (2) assess intra-rater reliability and agreement.

Method: Thirty participants recruited to the pilot trial were included in the analysis. The gWMFT was administered within two weeks of recruitment and at three months. Two occupational therapists scored participants through either direct observation or video. Reliability was assessed using intraclass correlation coefficients. Item-level agreement was assessed using proportion of agreement for functional ability scores and standard error of measurement for performance time. Total agreement was assessed using Bland-Altman plots and standard error of measurement.

Results: Excellent inter-rater reliability (n=28) was found between scoring through direct observation and by video (intraclass correlation coefficients >0.8) and excellent intra-rater reliability (n=21) was found (intraclass correlation coefficients >0.8), for item-level and summary scores. Low agreement was found between raters at item level. Adequate agreement was found for total functional ability, with increased measurement error found for scoring total performance time.

Conclusion: The gWMFT is a reliable measure of upper limb function and findings indicated video recording for scoring purposes may not be required. However inadequate agreement was found for performance time, indicating cautious application of the gWMFT. In view of low agreement, future studies should increase rater training and clarify functional ability scale ratings.

3.2 Chapter overview

The pilot RCT examined the effectiveness of mirror therapy with participants who were within three months of stroke onset. As part of the assessment of outcome measures, participants were video recorded completing the gWMFT at all time points. Using participant video recordings this chapter examined the reliability and agreement of the gWMFT at two assessment time points. This study was published in the British Journal of Occupational Therapy (Turtle *et al*. 2020) (Appendix 6). The researchers involved in this study and their roles are described in Appendix 1.

3.3 Introduction

Chapter 2 consisted of a systematic literature review which examined how the gWMFT has been implemented and reported in upper limb effectiveness studies and included a review of its psychometric properties. The 12 included studies were generally of a low quality and omitted essential information regarding test implementation and scoring. There were only two studies which assessed the reliability of the gWMFT in the chronic stage of stroke, with different versions and language applied, and scored ‘poor’ and ‘doubtful’ (Turtle *et al*. 2019). Chapter 2 highlighted the necessity of examining the psychometric properties of the gWFMT to support its use in clinical trials.

As discussed in Chapter 1 assessment of upper limb function is essential in capturing the effectiveness of upper limb interventions following stroke. The use of activity-level outcome measures provide valuable information regarding how an individual may use their upper limb in day to day activities and are considered essential tools for occupational therapists (Rowland and Gustafsson 2006). The WMFT is a widely used

and recommended activity-level outcome measure (Murphy *et al*. 2015; Santisteban *et al*. 2016), however its ability to capture the functional ability of individuals with severe upper limb impairment and those in the sub-acute phase of stroke is limited (Thompson-Butel *et al*. 2015; Lin *et al*. 2009a). The gWMFT, developed for individuals with moderate to severe hemiplegia, provides an alternative to the WMFT that could be used in the early stages of stroke (Constraint-Induced Movement Therapy Research Group 2002).

As detailed in Chapter 2 and by Turtle *et al*. (2019), intra-rater reliability of the 14-item gWMFT (Constraint-Induced Movement Therapy Research Group 2000) was assessed by Bonifer *et al*. (2005), as part of a constraint-induced movement therapy intervention study. Bonifer *et al*. (2005) found a high level of intra-rater reliability (Pearson's product moment correlation, r =0.96) for scoring functional ability in 20 individuals more than 12 months post-stroke. Pereira *et al*. (2015) completed an inter-rater reliability and agreement study of the Brazilian Portuguese version of the 13-item gWMFT (Constraint-Induced Movement Therapy Research Group 2002), which examined both functional ability and performance time. However, there were issues with how both studies were completed with inadequate detail reported regarding aspects of scoring and method of analysis (Bonifer *et al*. 2005; Pereira *et al.* 2015).

Examination of reliability in this study is based on classical test theory. Classical test theory is based on the premise that the observed score of individuals across subjective outcome measures is composed of the true score plus a random error score (DeVellis 2006). The true score is the average of an individual's score if they were to complete the test an infinite number of times (DeVellis 2006). The examination of reliability aims to determine the degree to which an outcome measure is free from random error.

However, the examination of psychometric properties using classical test theory means that results are dependent on the study sample, and as such is a property of those scores (Streiner 2010).

In this study reliability, agreement and internal consistency were examined. The assessment of participants by different assessors and over multiple time-points requires consistent scoring to allow for meaningful interpretation and comparison across results and is examined through inter- and intra-rater reliability. Agreement was also considered between assessors (inter-rater) and over multiple time-points (intra-rater). Agreement assesses the degree of measurement error and is considered a property of the outcome measure due to examination of how identical scores are regardless of who scores participants and when. The guidelines for reporting reliability and agreement studies advise reporting both measurement properties (Kottner *et al*. 2011). Essential to classical test theory is that test items demonstrate high internal consistency, test items which correlate well with each other indicate the items are measuring the same construct (Streiner 2003). Internal consistency relies on a single administration of an outcome measure and is the most widely reported element of reliability (Streiner 2003).

Clinical utility is related to how useful an outcome measure is in clinical practice incorporating factors related to how practical, accessible, acceptable and appropriate the outcome measure is for both clinicians and patients (Smart 2006). The gWMFT was developed as a laboratory-based outcome measure and as such may have limited clinical utility. Furthermore, authors of the gWMFT recommend individuals are video recorded for the purpose of scoring the FAS. The WMFT can take up to 40 minutes to administer (Bogard *et al*. 2009) and scoring functional ability at a later time adds to the

burden of delivery. Whitall *et al*. (2006) advised the WMFT FAS could be scored during test administration in order to reduce rater burden, with high levels of inter-rater reliability found between raters scoring through direct observation and recorded videos.

There is no known evidence of the reliability and agreement of the gWMFT administered within three months of stroke onset. in order to support the use of the gWMFT in large scale trials there is a need to examine the reliability of the gWMFT when scored by multiple raters and to consider if video recording is necessary for scoring.

3.4 Aim and objectives

3.4.1 Aim

The aim of this study was to investigate the reliability and agreement of the gWMFT in a sub-acute stroke population.

3.4.2 Objectives

The objectives of the study were:

1. To investigate the inter-rater reliability and agreement of the gWMFT when used by rater scoring through direct observation and using recorded videos.
2. To investigate the intra-rater reliability and agreement of the gWMFT scoring using recorded videos.
3. To investigate the internal consistency of the gWMFT at two weeks and three months assessments.

3.5 Method

3.5.1 Study design

An inter- and intra-rater reliability and agreement study was completed as part of a pilot RCT examining the effectiveness of mirror therapy with sub-acute stroke participants. This study is presented based on the published guidelines for reporting reliability and agreement studies (Kottner *et al*. 2011).

3.5.2 Participants

Participants (n=30) consecutively recruited to the pilot RCT between May 2015 and March 2017 from three hospital sites in the NHSCT in Northern Ireland were included (ClinicialTrials.gov identifier: NCT02276729). Inclusion criteria were: adults aged 18 years plus and recently admitted to an inpatient rehabilitation ward; stroke diagnosis within three months with upper limb motor loss and upper limb rehabilitation a key component of treatment; able to understand and follow two-part verbal and written commands in the English language and able to provide written consent. Exclusion criteria were: previous stroke; gross cognitive impairment.

3.5.3 Ethical approval

Ethical approval was granted by the Office for Research and Ethics Committees Northern Ireland (Ref:14/NI/1149) and research governance was granted by the NHSCT prior to study commencement (Appendix 7 Ethical approval from the Office for Research Ethics Committees, Appendix 8 Study approval from NHSCT research governance). Written informed consent was obtained from participants, with additional written informed consent gathered for the recording of video footage for scoring purposes.

3.5.4 Raters

Rater one and rater two were research occupational therapists. The therapists were employed solely to collect outcome measures on the trial and had no clinical relationship with the participants. Training for both raters involved reviewing the manual (Constraint Induced Movement Therapy Research Group, 2002) and viewing training videos, the scoring of which was verified by occupational therapists experienced in the clinical administration of the outcome.

3.5.5 Outcome measure

The gWMFT assesses timed performance and quality of movement (Constraint Induced Movement Therapy Research Group, 2002) and is described in detail in Chapter 2. The gWMFT consists of 13 graded test items (Appendix 3) (Constraint Induced Movement Therapy Research Group, 2002) and takes approximately 40 minutes to administer. Video recording of the gWMFT is recommended to enable retrospective scoring of functional ability.

Scoring of the gWMFT

Quality of movement is assessed on the gWMFT using a FAS. This is an eight-point ordinal scale, ranging from zero (not attempted) to seven (normal movement). Items are completed on two levels (A and B), where level A items are of a higher level of difficulty and are scored between four and seven. Level B items are of a lower level of difficulty and are scored between zero and three. Any item not completed are scored zero. For the assessment of performance time, participants have 30 seconds to complete level A items, and if unable to do so have a second opportunity to complete the task at level B. Sixty seconds are added onto performance time for level B items,

with a maximum time of 120 seconds. The scoring procedure for level A and level B test items is presented in Appendix 4.

3.5.6 Procedure

The gWMFT was administered and video recorded according to protocol guidelines by one occupational therapist (rater one) (Constraint Induced Movement Therapy Research Group, 2002). To standardise placement of objects and participants, a template was devised from a plexiglass sheet according to protocol instructions and securely affixed to a table-top (Appendix 9). Participants were seated at the table on a wheelchair to minimise disruption when adjusting position. The gWMFT was used to assess the participants affected arm.

As part of the pilot RCT participants were assessed at baseline and every two weeks until discharge, and at three- and six-month follow-up. Assessments completed at two weeks (T1) and three months (T2) were included in this analysis. The assessments completed at T1 took place in a private room used for research purposes on the hospital site. Assessments completed at T2 generally took place in the participant's own home. All recorded participant footage was viewed in a private room on hospital premises. Raters were blinded to each other's scoring.

For inter-rater analyses, rater one completed scoring through direct observation and rater two later viewed and scored participant videos for assessments completed at T1. For intra-rater analyses, rater two scored assessment videos completed at T2 and re-scored one month later. Internal consistency was assessed for both raters at T1 and for both testing sessions at T2.

3.5.7 Data analysis

Descriptive statistics for age, gender and side of hemiparesis were recorded. The mean value was reported for the total FAS score, and the median value was reported for total performance time (Constraint Induced Movement Therapy Research Group 2002). Score distributions were examined for both time points. Floor and ceiling effects were present if 15% or more of the sample achieved the minimum or maximum scores (McHorney and Tarlov 1995).

Item-level reliability and agreement were completed to determine if there were any issues with individual items of the gWMFT. Inter-rater reliability for total and item-level functional ability and performance time were assessed using a two-way random, consistency ICC ($ICC_{2,1}$) (Shrout and Fleiss 1979). This was used because all participants were scored by two raters independently and enables generalisations to be made to other raters within the same population.

Intra-rater reliability for total and item-level functional ability and performance time were assessed using two-way mixed effects, consistency ICC ($ICC_{3,1}$) (Shrout and Fleiss 1979). Intraclass correlation coefficients determine the level of consistency in the ranking of scores (Hallgren 2012). A reliability score of 0.60 and above was considered acceptable (Cicchetti 1994).

To examine item-level inter- and intra-rater agreement, proportion of agreement and proportion of agreement ± 1 point were completed for functional ability. Standard error of measurement (SEM) (Stratford and Goldsmith 1997) was completed for item-level performance time. Bland and Altman plots and SEM were calculated for the total scores of both functional ability and performance time. Using a scatter graph Bland-

Altman plots display the difference between rater scores against the mean and illustrate the 95% limits of agreement, which are the mean difference of scores ± 1.96 × standard deviation (SD) (Bland and Altman 1986). Narrow limits of agreement indicate increased agreement between raters. The SEM was calculated from the square root of the mean square error (De Vet *et al*. 2006; Stratford and Goldsmith 1997). Expressed in the original units of measurement, the SEM portrays the amount of measurement error in scoring; the larger the value, the greater the variability between raters.

Internal consistency of functional ability and performance time were analysed using Cronbach's alpha. Internal consistency is indicative of the interrelatedness of individual test items and does not require multiple test sessions or raters (Streiner 2003). Values above 0.70 were considered acceptable (Terwee *et al*., 2007).

Bland-Altman plots were completed using Microsoft Excel 2016. Data analysis was completed using SPSS Version 25.0 (SPSS Inc. Chicago) and Microsoft Excel 2016.

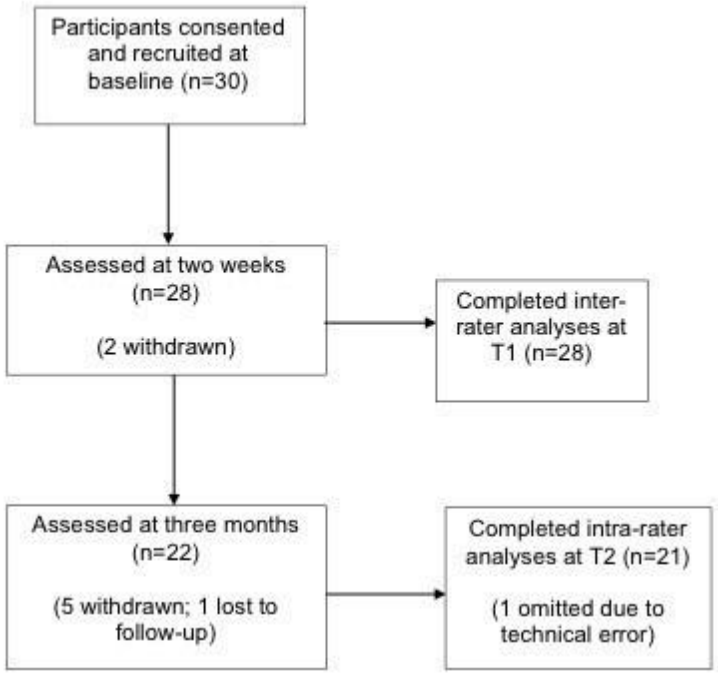
3.6 Results

A total of 30 participants were recruited (mean days post-stroke (SD), 14.73 (SD=8.36)). At T1 two participants were withdrawn from the study for medical reasons. Technical errors were identified in the recordings for two participants at T1. In order to utilise existing data, summary scores were calculated using the available items for both participants. Rater one and rater two scored 28 participants at T1.

At T2 five participants were withdrawn for medical reasons and one participant was lost to follow up. In addition, rater one was unable to complete a participant's

assessment at three months, due to lack of space in the participant's home for the testing kit. Twenty-two participants were assessed by rater one at T2. Technical errors were identified in the recordings for three participants at three months. Summary scores were calculated using the available items for two participants. However, due to numerous recording errors data from one participant was excluded from intra-rater analyses. As a result, rater two scored 21 participants at T2 using video. Figure 3.1 demonstrates participant flow through the study.

Figure 3.1 Flowchart of study participants



Data from 28 participants were included for inter-rater analyses (mean age (SD), 71.3 (9.85); 18 males and 10 females) and data from 21 participants were included for intra-

rater analyses (mean age (SD), 70.5 (8.7); 16 males and five females). Participant characteristics are presented in Table 3.1.

Table 3.1 Participant characteristics and graded Wolf Motor Function Test scores

	Two Weeks (T1) (n=28)		**Three Months (T2) (n=21)**	
Gender				
Male, n	18		16	
Female, n	10		5	
Age in years, mean (SD)	71.3 (9.6)		70.5 (8.7)	
Side of hemiplegia				
Left, n	18		15	
Right, n	10		6	
gWMFT FAS	**Rater one**	**Rater two**	**Session one**	**Session two**
Mean (SD)	3.74 (2.47)	3.16 (2.11)	3.45 (2.28)	3.53 (2.35)
Floor effect, n (%)	6 (21.4%)	6 (21.4%)	4 (19%)	4 (19%)
Ceiling effect, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
gWMFT performance time	**Rater one**	**Rater two**	**Session one**	**Session two**
Mean (SD)	51.79 (55.18)	53.94 (54.51)	47.74 (55.74)	46.39 (55.77)
Floor effect, n (%)	10 (35.7%)	10 (35.7%)	7 (33.3%)	7 (33.3%)
Ceiling effect, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Abbreviations: SD, standard deviation; gWMFT, graded Wolf Motor Function Test; FAS, functional ability scale; %, percentage

3.6.1 Floor and ceiling effects

Ceiling effects were not evident for either assessment session. At T1, floor effects were found for performance time and functional ability by both raters, with 35.7% and 21.4% of the sample achieving the maximum score of 120 seconds and minimum score of zero, respectively (Table 3.1).

At T2, floor effects were found for performance time, with 33.3% of the sample achieving the maximum score of 120 seconds (Table 3.1). Floor effects were also found for functional ability at both testing sessions, with 19% of the sample achieving the minimum score of zero (Table 3.1).

3.6.2 Inter-rater reliability and agreement

High levels of reliability were found between rater one scoring through direct observation and rater two scoring using recorded videos for item-level (Table 3.2) and total (Table 3.3) functional ability and performance time, with ICC values above 0.8.

The proportion of agreement for scoring functional ability at item-level ranged from 0.43 to 0.64 and proportion of agreement ± 1 ranged from 0.56 to 0.96 (Table 3.2). Agreement based on SEM values for performance time at item-level ranged from 0.32 to 19.30, with greater differences found for scoring items one and four through to twelve (Table 3.2). Standard error of measurement values for total scores were 0.33 for functional ability, and 6.49 for performance time (Table 3.3). Larger differences for scoring performance time occurred where there were differences between raters in assigning participant performance to level A or level B tasks

Table 3.2 Item-level reliability and agreement for the graded Wolf Motor Function Test

	Inter-rater					**Intra-rater**				
	Reliability $ICC_{(2,1)}$ (95%CI)		**Agreement**			**Reliability $ICC_{(3,1)}$ (95%CI)**		**Agreement**		
	FAS	**Time**	**FAS**		**Time SEM**	**FAS**	**Time**	**FAS**		**Time SEM**
			Po	**Po±1**				**Po**	**Po±1**	
1- Raise forearm to table (side)	0.884 (0.764-0.944)	0.943 (0.880-0.973)	0.43	0.82	11.11	0.976 (0.940-0.990)	1 (1-1)	0.8	1	0.56
2- Raise forearm from table to box (side)	0.967 (0.930-0.985)	1 (1-1)	0.64	0.93	0.32	0.982 (0.956-0.993)	1 (1-1)	0.81	1	0.54
3- Extend elbow (side)	0.967 (0.931-0.985)	1 (1-1)	0.54	0.96	0.70	0.969 (0.926-0.987)	0.970 (0.929-0.988)	0.71	0.95	9.29
4- Extend elbow against 1 lb. weight (side)	0.866 (0.728-0.937)	0.853 (0.704-0.930)	0.44	0.56	19.30	0.948 (0.875-0.978)	0.967 (0.919-0.986)	0.81	0.95	9.25
5- Raise hand to table (front)	0.913 (0.820-0.959)	0.923 (0.841-0.964)	0.43	0.86	13.22	0.926 (0.827-0.969)	0.969 (0.924-0.987)	0.62	0.95	9.27
6- Raise hand to box (front)	0.926 (0.846-0.965)	0.969 (0.934-0.985)	0.57	0.86	9.89	0.988 (0.970-0.995)	1 (1-1)	0.86	1	0.20
7- Reach and retrieve 1 lb. weight on table	0.953 (0.902-0.978)	0.919 (0.833-0.962)	0.57	0.86	13.65	0.967 (0.920-0.986)	1 (1-1)	0.57	1	0.10
8- Move foam stick through supination and pronation	0.900 (0.797-0.953)	0.901 (0.797-0.953)	0.43	0.93	17.16	0.984 (0.960-0.993)	1 (1-1)	0.76	1	0.07
9- Grasp and lift washcloth	0.946 (0.888-0.975)	0.939 (0.872-0.971)	0.5	0.89	13.43	0.973 (0.936-0.989)	0.972 (0.933-0.989)	0.67	0.90	9.24
10- Flip light switch	0.932 (0.858-0.968)	0.936 (0.867-0.970)	0.57	0.86	13.34	0.976 (0.941-0.990)	1 (1-1)	0.62	1	0.08

11- Grasp and lift pen	0.902 (0.797-0.954)	0.875 (0.745-0.941)	0.52	0.85	17.98	0.954 (0.890-0.981)	0.969 (0.924-0.987)	0.77	0.95	9.24
12- Grasp and lift cotton balls	0.913 (0.820-0.959)	0.914 (0.823-0.959)	0.57	0.82	15.08	0.953 (0.888-0.981)	0.957 (0.898-0.983)	0.86	0.95	9.25
13- Lift weighted basket (3 lb.), place onto raised table (standing)	0.971 (0.939-0.987)	1 (1-1)	0.68	0.93	0.54	0.987 (0.968-0.995)	1 (1-1)	0.8	1	0.08

Abbreviations: ICC, intraclass correlation coefficient; FAS, functional ability scale; Po, proportion of observed agreement; Po±1, proportion of agreement ±1 point; SEM, standard error of measurement

Table 3.3 Inter- and intra-rater reliability, standard error of measurement and internal consistency of the graded Wolf Motor Function Test

	Inter-rater reliability $ICC_{2,1}$ (95% CI) (n=28)	**Intra-rater reliability $ICC_{3,1}$ (95% CI) (n=21)**	**SEM**		**Internal Consistency**			
			Inter-rater (n=28)	**Intra-rater (n=21)**	**Two Weeks (n=28)**		**Three Months (n=19*)**	
					Rater 1 (n=28)	**Rater 2 (n=26)**	**Session 1**	**Session 2**
Functional ability	0.979 (0.955-0.990)	0.993 (0.983-0.997)	0.33	0.19	0.99	0.98	0.99	0.98
Performance time	0.986 (0.970-0.993)	0.996 (0.990-0.998)	6.49	3.64	0.98	0.98	0.98	0.99

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; SEM, standard error of measurement.

*Due to technical difficulties one item was not scored for participants one and two, leading to their exclusion as part of the internal consistency analysis.

Visual inspection of the scatterplot for functional ability (Figure 3.1) highlighted possible proportional bias, where greater differences were seen between raters as the mean score increased. A simple regression analysis confirmed mean functional ability significantly predicted differences in rater scoring (p<0.05), R^2 value was 0.575 (Figure 3.2). Proportional bias reduces the accuracy of the limits of agreement, and a Bland-Altman plot based on the ratio between rater scores against their mean was constructed (Figure 3.2) (Bland and Altman 1986; Bland and Altman 1999). One participant was removed, as a ratio value cannot be calculated with a score of zero (O'Donoghue 2012). There was no evidence of proportional bias for the ratio Bland-Altman plot (p>0.05). The limits of agreement were between 0.73 and 1.63, indicating adequate agreement, with one outlier above the limits of agreement (see Figure 3.2).

Figure 3.2 Scatterplot showing the relationship between the difference in rater scores for functional ability at two weeks, including line of regression

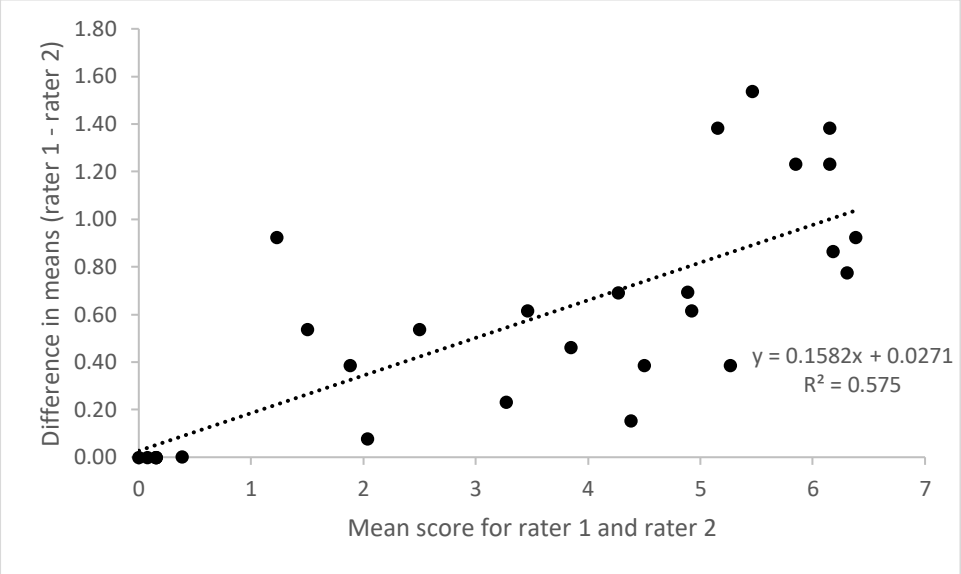
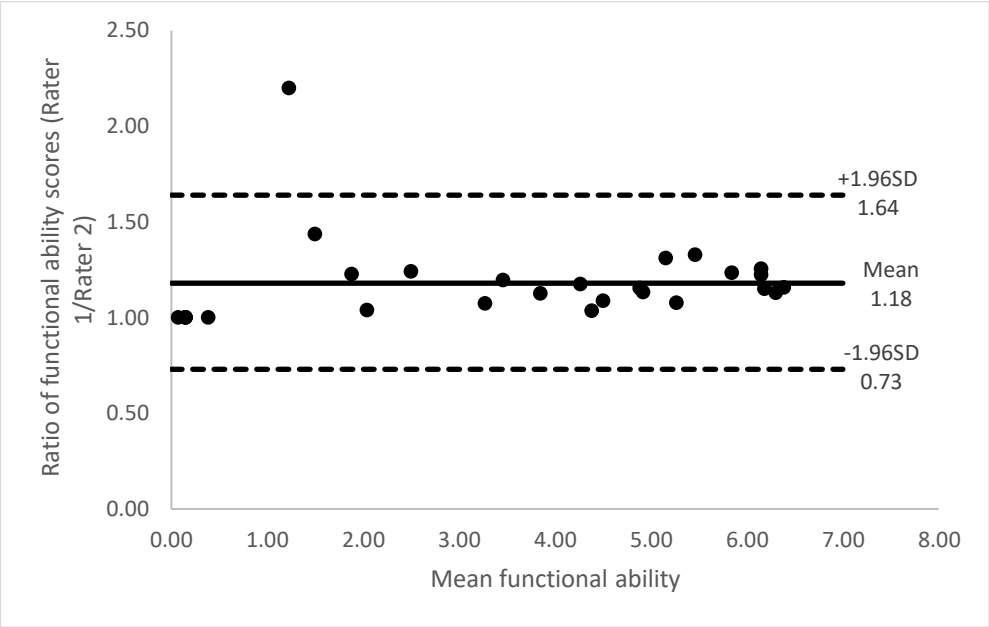
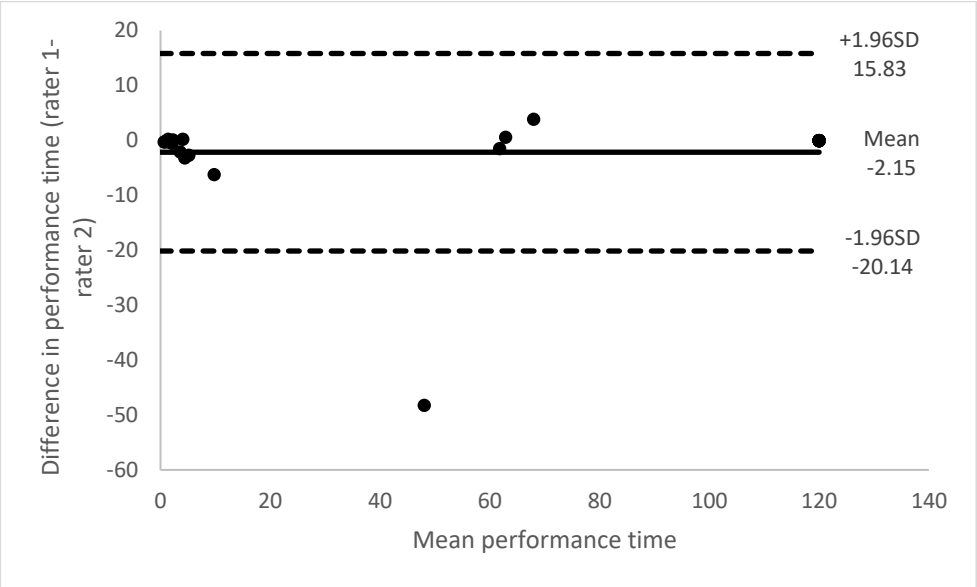


Figure 3.3 Bland-Altman plot for inter-rater agreement of functional ability using the ratio scores for rater one and rater two



The limits of agreement for performance time were between -20.14 and 15.83 (see Figure 3.3). An outlier below the limits of agreement presented a difference of 48.22 seconds between raters. Larger differences for scoring performance time occurred where there were differences between raters in assigning participant performance to level A or level B tasks.

Figure 3.4 Bland-Altman plot for inter-rater agreement for performance time using differences between rater 1 and rater 2 plotted against their average



3.6.3 Intra-rater reliability and agreement

High levels of reliability were found for item-level (Table 3.2) and total (Table 3.3) functional ability and performance time, with ICCs above 0.9. Proportion of agreement ranged from 0.57 to 0.86 and proportion of agreement ± 1 ranged from 0.90 to 1 for functional ability scores at item-level (Table 3.3). Agreement based on SEM values for item-level performance time ranged from 0.07 to 9.29, with greater differences found for scoring items three, four, five, nine, eleven and twelve (Table 3.2). Standard error of measurement values for total scores were 0.19 for functional ability, and 3.64 for performance time (Table 3.3).

The limits of agreement for functional ability ranged from -0.6 to 0.47, indicating variability within one point between the two scoring sessions (Figure 3.4). The limits of agreement for performance time were between -8.73 and 11.43 (Figure 3.5). An outlier above the limits of agreement presented a difference of 23.65 seconds between scoring sessions for one participant. Removal of this outlier led to a reduction in the

mean difference to 0.23 and the limits of agreement were between -0.97 and 1.43. Differences between raters occurred where there were disagreements in assigning participant performance to level A or level B tasks.

Figure 3.5 Bland-Altman plots for intra-rater agreement for functional ability using differences between testing sessions plotted against their average

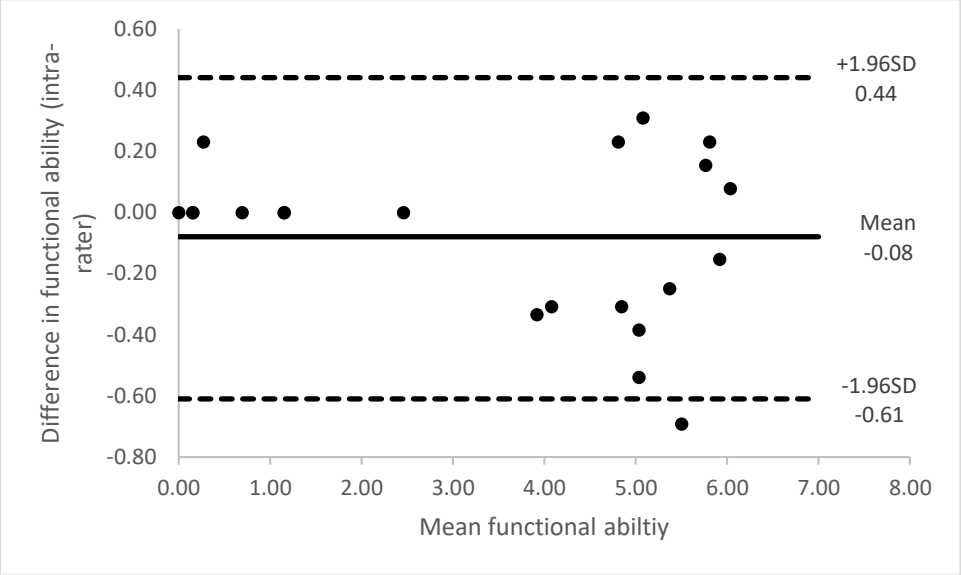
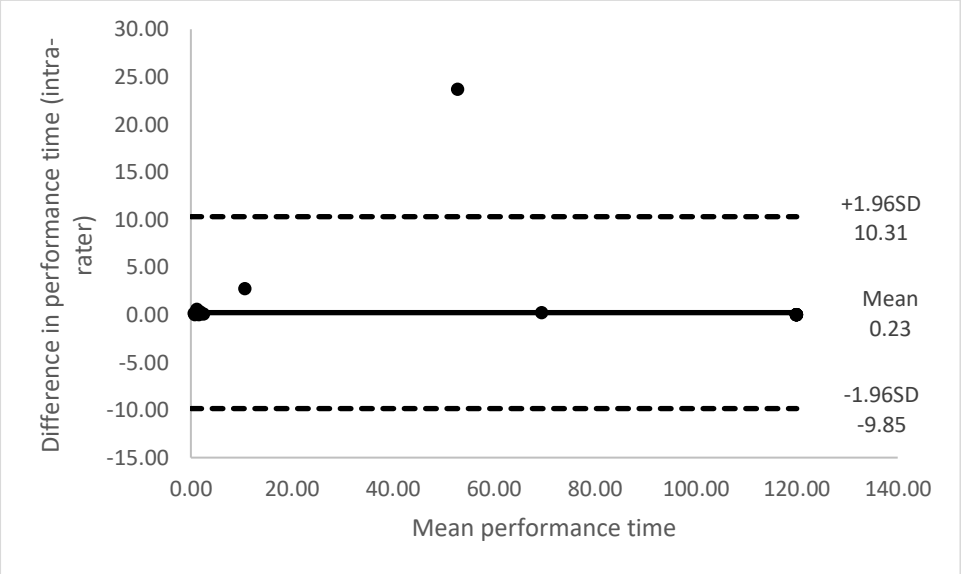


Figure 3.6 Bland-Altman plots for intra-rater agreement for performance time using differences between testing sessions plotted against their average



3.6.4 Internal consistency

Internal consistency values for functional ability and performance time for all assessment points were above 0.9 (Table 3.3). Inter-item correlations for functional ability were good (>0.7) and adequate for performance time (>0.4).

3.7 Discussion

The aim of this study was to examine the reliability and agreement of the gWMFT with a sub-acute stroke cohort, as part of the feasibility aims of a pilot RCT. Results between raters scoring through direct observation and using video were compared, and excellent inter-rater reliability was found for scoring functional ability and performance time, with adequate agreement found for scoring functional ability. Similar results were found for intra-rater reliability and agreement analyses. However, an unacceptable degree of measurement error was found for performance time for both inter- and intra-rater agreement analyses. This is the first known study to examine the reliability and agreement of the gWMFT in the early stages of stroke.

Floor effects were found for performance time and functional ability, with more than 30% of participants achieving the maximum score of 120 seconds and more than 15% of the sample achieving the minimum sore of zero on the FAS. In the study by Pereira *et al*. (2015) floor and ceiling effects were not reported for the Brazilian-Portuguese version of the gWMFT. However, Pereira *et al*. (2015) reported mean performance time scores, which are less likely to convey floor effects in comparison to reporting median performance time (Whitall *et al*. 2006; Hodics *et al*. 2012). Furthermore Pereira *et al.* (2015) completed this study with individuals more than six months post-stroke which may account for why floor effects were not found. Examination of the distribution of scores for the WMFT by Thompson-Butel *et al*. (2015) found that out of 16

participants who were categorised with severe upper limb impairment, five participants were unable to complete any test item within 120 seconds. Similarly, Lin *et al*. (2009a) found floor effects for WMFT functional ability scores when delivered within 14 days of stroke onset. As part of a pragmatic clinical trial, participants in the current study were not required to have a minimum level of function and as such presented with varying levels of upper limb ability. Consequently, the gWMFT was not able to sensitively measure the upper limb capabilities of all participants who were recruited, and it was not clear to which level of upper limb ability the gWMFT is best applied.

There were high levels of inter-rater reliability found between raters scoring through direct observation and by video which indicated that scoring by video may not be a necessary adjunct. This was further substantiated by adequate agreement found between raters scoring functional ability. Whitall *et al*. (2006) investigated scoring the WMFT by video and through direct observation by assessing the reliability between three raters who scored participants using recorded videos and one rater who scored participants through direct observation. Excellent levels of reliability with ICCs greater than 0.9 were reported for functional ability and performance time which indicated scoring by video may not be necessary (Whitall *et al*. 2006). Pereira *et al*. (2015) reported similar levels of inter-rater reliability for total and item-level functional ability of the gWMFT with two raters who independently scored participants using recorded performances. However, for total and item-level performance time, the confidence intervals reported by Pereira *et al*. (2015) for six items were wide indicating the true ICC value could be evident of poor reliability (Koo and Li 2016).

Agreement for total functional ability was adequate with poor exact agreement across all items. However, this was generally due to differences of one point between raters,

and as such were considered acceptable. Adequate agreement was found for inter-rater agreement for total functional ability (SEM=0.33). Similarly, Pereira *et al*. (2015) reported acceptable agreement between raters scoring total functional ability using Bland and Altman plots with a mean difference of -0.04 between raters.

In the current study the Bland-Altman plot for inter-rater agreement of functional ability at two weeks highlighted proportional bias, where the mean difference between raters increased as scores on the gWMFT FAS increased. This could be the result of difficulty differentiating between participants with higher levels of functional ability. Although, recommended by authors of the gWMFT and the WMFT (Constraint Induced Movement Therapy Research Group 2002; Taub *et al*. 2011), the least affected limb was not tested. Scores for the less affected limb may act as a comparison for the more affected limb and help raters discern between FAS ratings accordingly.

Inadequate agreement was found between raters scoring performance time at item-level and for total scores. The SEM for performance time highlighted greater discrepancies between raters. Examination of scores at item level highlighted rater variations in assigning participant performance to level A or level B, with SEM values greater than nine seconds found for 10 items. The Bland-Altman plots confirmed this finding, with wide limits of agreement and outliers, resulting in skewed outcomes. Although, Pereira *et al*. (2015) modified the gWMFT FAS and enhanced details to identify compensatory movements and abnormal synergies, a greater degree of measurement error was observed in the Bland and Altman plots for scoring performance time. The limits of agreement were between -0.68 and 16.1, indicating disagreements between raters (Pereira *et al*. 2015).

Whilst the raters underwent training separately, the training content was consistent for both. This comprised reading the manual (Constraint Induced movement Therapy Research Group 2002), viewing training videos of an experienced occupational therapist administering the test with stroke survivors and scoring in real time. This was augmented by a review of the scoring results with an experienced occupational therapist in a training session. In previous studies raters have been required to demonstrate approximate scoring to each other prior to study commencement (Morris *et al*., 2001; Whitall *et al*., 2006). This was not required in the current study, potentially leading to measurement error and the disagreements demonstrated at item-level.

Excellent intra-rater reliability for functional ability and performance time were found indicating consistent scoring by one rater, over a one-month interval. Intra-rater SEM values for functional ability displayed minimal variation between scoring sessions, indicating a good level of agreement. However, similar to inter-rater agreement analyses, there were unacceptable differences in scoring performance time. This was also reflected by wide limits of agreement in the Bland-Altman plot, with an outlier representing a difference of 23.64 seconds between scoring sessions. Higher levels of intra-rater reliability and agreement were found in comparison to inter-rater analyses indicating that scoring by one rater produces more consistent results. This finding creates uncertainty for the consistent delivery of the gWMFT in a main RCT.

Whilst differences in scoring modality may have impacted on rater differences in the current study, unacceptable measurement error was also found for scoring performance time using video alone, as illustrated by intra-rater analyses. Similarly, Pereira *et al*. (2015) found wide limits of agreement when scoring using video alone. This indicates the presence of additional factors impacting on measurement error. This

is likely the result of differences in accurately differentiating between a level A and level B performance by participants. Cramer *et al*. (2017) reported on the specific need to standardise the scoring of outcome measures across clinical trials to improve stroke research. Duff *et al*. (2015) recognised the issues of variability in ascribing the subjective aspects of the WMFT to patient performance and designed a quality process to ensure rater standardisation. This included clarification of the WMFT FAS, with essential elements added to each test item to indicate when a task has been successfully completed with photographs to illustrate (Duff *et al*. 2015). The adoption of a standardised procedure for administration and scoring of the Fugl-Meyer upper limb assessment led to increased accuracy and reduced variance in scoring (See *et al*. 2013). Therefore, greater clarity on scoring using the FAS is recommended in future studies to enable accurate differentiation between items completed at either level A or level B is recommended.

A high level of internal consistency was found for functional ability scores and performance time in the current study, supporting the use of the gWMFT early following stroke and at three months. This is the first known study to assess the internal consistency of the gWMFT with sub-acute stroke survivors. Edwards *et al*. (2012) found similar levels of internal consistency for the WMFT when administered to stroke survivors within 28 days of stroke onset and at three months post-stroke.

3.7.1 Limitations

This study was developed as part of an ongoing pilot RCT and as such the sample size was small. In addition, this study examined stroke survivors in the sub-acute phase of stroke and most experienced difficulty attempting all test items. Therefore, consideration of reliability and agreement estimates should be applied with caution.

Future study with a larger sample size could stratify participants according to level of ability and examine use of the gWMFT in chronic stroke. This would enable exploration of the gWMFT to determine which levels of ability it can more sensitively measure.

Designed as a laboratory-based outcome, rater one found difficulty with application of the gWMFT in a community-based setting. One participant was unable to complete a three-month assessment, due to lack of space in the home. This indicates there are limits to the measure's ease of use beyond a clinical setting, with implications for research and longitudinal follow-up.

3.8 Conclusion

The reliability and agreement of the gWMFT was examined with participants in the sub-acute phase of stroke as part of the feasibility aims of a pilot RCT. High levels of inter- and intra-rater reliability and agreement were found for both domains of the gWMFT, with acceptable agreement found for functional ability scored by two raters and over two testing sessions. In addition, results indicated that scoring the gWMFT by video may not be necessary, supporting its use in clinical practice. However, agreement analyses highlighted greater measurement error for scoring performance time. As such, further research is needed to examine the impact of standardised, supplemental training on scoring the gWMFT to support its use in a main clinical trial.

Chapter 4

Chapter 4 – Responsiveness of the Functional Independence Measure and the graded Wolf Motor Function Test

4.1 Abstract

Background: The Functional Independence Measure (FIM) and the graded Wolf Motor Function Test (gWMFT) are activity-level outcome measures used in the evaluation of mirror therapy in the pilot trial. Responsiveness has not been assessed in the gWMFT, with limited assessment of external responsiveness of the FIM. Therefore, the aim of this chapter was to assess the responsiveness of the FIM and gWMFT in individuals with stroke.

Methods: Forty participants recruited to the randomised trial were scored on the motor, cognitive and total FIM, and on functional ability and performance time of the gWMFT. Internal responsiveness was assessed using effect size statistics; effect size and standardised response mean (SRM). The external criterion was patient-reported rating of change; correlations with change scores between baseline and discharge were used to determine the strength of association. Receiver operating characteristic (ROC) curves were completed where correlations were greater than 0.3.

Results: Marked ceiling effects were found for FIM cognitive scores and gWMFT performance time across all time intervals. Twenty-five participants completed the FIM at all four assessments points and 23 participants completed the gWMFT assessments at all four assessments points. The FIM total and motor scores demonstrated adequate internal responsiveness across all time intervals and groups. Negligible values of internal responsiveness were found

for FIM cognitive scores. Internal responsiveness of the gWMFT was greatest between baseline and discharge. Greater levels of responsiveness were found for the gWMFT functional ability in comparison to performance time. An acceptable correlation was found between the external criterion and gWMFT FAS, with an adequate area under the ROC curve score of 0.74.

Conclusions: The gWMFT, FIM motor and FIM total scores demonstrated adequate responsiveness. The gWMFT functional ability was sensitive to patient-perceived changes in upper limb function.

4.2 Chapter overview

This chapter was undertaken to examine the responsiveness of the activity-level outcome measures which were included in the pilot RCT. This chapter sought to determine the suitability of the outcome measures and time intervals chosen for assessment for use in a main trial. Appendix 1 details the team members involved in the study.

4.3 Introduction

Chapter 3 examined the reliability and agreement of the gWMFT within three months of stroke onset and was the first known study to do so. Overall, high levels of reliability were found for the gWMFT. As part of the feasibility aims of a pilot RCT, the gWMFT was considered suitable for use in a main RCT. However, agreement analyses highlighted discrepancies between raters as well as scoring differences over time. The development of a standardised training programme implemented before the commencement of a main trial and supplemented throughout trial delivery to improve accuracy in scoring and reduce measurement error is recommended. In addition, further examination of the psychometric properties of the gWMFT was warranted.

Measurement of function during stroke rehabilitation also requires outcome measures which are responsive to change (Cohen and Marino 2000). Responsiveness refers to the ability of an outcome measure to detect clinically relevant change over time in the construct of interest (Mokkink *et al*. 2010). Adequate responsiveness is integral to determining the effectiveness of a treatment and important for calculating statistical power and sample size (Van

Der Putten *et al*. 1999; Hobart *et al*. 2007). There remains debate as to whether responsiveness is an aspect of validity or a psychometric property in its own right, with differing opinions on how to define and analyse responsiveness (Husted *et al*. 2000; Beaton 2000; Mokkink *et al*. 2010). Husted *et al*. (2000) reported two approaches to responsiveness; internal and external responsiveness. Internal responsiveness is evaluated by examining the changes in scores across a specified timeframe and is also referred to as the distribution-based method of responsiveness (Revicki *et al*. 2006). Effect size statistics such as Cohen's effect size (ES), standardised response mean (SRM) and Guyatt's responsiveness index are commonly used to determine the degree of change (Husted *et al*. 2000). The resulting scores are dependent on the intervention under investigation and the outcome measure used to illustrate that change.

External responsiveness, or the anchor-based method of responsiveness, describes the ability of an outcome measure to identify change in correspondence with an external criterion, indicative of meaningful change (Husted *et al*. 2000). This external criterion can be an outcome measure which is considered an accepted reference standard for change in the area of interest or information reported by patients or clinicians indicating whether they perceive change to have occurred or not (Husted *et al*. 2000). Recommended methods of analysis for external responsiveness include the use of receiver operating characteristic curves, correlations and regression analyses (Husted *et al*. 2000).

Upper limb impairment is one of the most commonly reported consequences of stroke (Lawrence *et al*. 2001; Gillen and Nilsen 2016), with motor impairment

experienced by 80% of survivors of stroke (Langhorne *et al*. 2009). Using the ICF (WHO 2002), the FIM and gWMFT are categorised as activity-level outcome measures (Sivan *et al*. 2011). The FIM (Uniform Data System for Medical Rehabilitation 1997) and the gWMFT (Constraint Induced Movement Therapy Research Group 2002) provide differential insights into upper limb activity following stroke.

The FIM is widely used in stroke rehabilitation settings to capture change according to the level of assistance required to complete ADLs (Bottemiller *et al*. 2006). Upper limb impairment is linked to greater difficulty completing ADLs, and the FIM is commonly used in the evaluation of upper limb treatments (Sivan *et al*. 2011; Cantero-Telléz *et al*. 2019). The FIM has been studied extensively in terms of its psychometric properties (Ottenbacher *et al*. 1996; Stineman *et al*. 1996; Van der Putten *et al*. 1999). Using meta-analysis, high levels of inter-rater and test-retest reliability were found for the FIM across 11 studies with 1568 participants who presented with a range of conditions, including stroke. Examination of the responsiveness of the FIM has demonstrated large effect sizes when delivered to stroke survivors before and after treatment at a rehabilitation centre (Van der Putten *et al*. 1999), with similar results found for the FIM motor sub-scale (Cano *et al*. 2006). Hobart *et al*. (2001) examined the responsiveness of the FIM with individuals who ranged from 3 weeks post-stroke up to 12 years post stroke (n=45), with 27 participants recruited within three months of stroke onset. Standardised response means (SRM) were similar for the FIM total (SRM = 0.48) and FIM motor sub-scale (SRM = 0.54), with a much lower level of responsiveness found for the FIM cognitive sub-scale (SRM = 0.17). Ceiling effects and low levels of responsiveness have been found

for FIM cognitive scores when used up to six-months post-stroke (Schepers *et al*. 2006), which may limit the utility of the FIM cognitive sub-scale in research and clinical settings (Ward *et al*. 2014). One study examined the minimal clinically important difference (MCID) for the FIM using clinician perceptions of recovery following stroke (Beninato *et al*. 2006), with studies also using the Modified Rankin Scale as an external criterion (Wallace *et al*. 2002). However, the psychometric properties of outcome measures are not fixed, varying with the sample group, the study context and the intervention under investigation (Liang *et al*. 1990; Husted *et al*. 2000). In addition, there has been no examination of the responsiveness of the FIM in relation to patient-perceived recovery, which is viewed as a significant element of responsiveness, constituting actual meaningful change (Revicki *et al*. 2008).

The gWMFT which was examined extensively in Chapter 2, was designed to capture upper limb function for individuals with hemiplegia following stroke or traumatic brain injury (Constraint induced Movement Therapy Research Group 2002). A high level of inter- and intra-rater reliability was reported for the gWMFT when used within three months of stroke onset (Turtle *et al*. 2020). However, psychometric evaluation of the gWMFT remains limited with no examination of the responsiveness of the gWMFT.

Advocated by guidelines published by the Medical Research Council, pilot trials provide an opportunity to examine any issues before implementing a large-scale RCT and that includes exploring how appropriate an outcome measure is for the study setting (Craig *et al*. 2008; Jones 2018). Therefore, the responsiveness of the activity-level outcome measures was assessed to

examine the ability of the outcome measures to capture change across the assessment periods included in the study and to examine the relevance of those time points.

4.4 Aim and objectives

4.4.1 Aim

The aim of this study was to examine the activity-level outcome measures completed as part of the pilot RCT.

4.4.2 Objectives

The objectives of the study were:

1. To assess the internal and external responsiveness of the gWMFT between baseline and discharge; baseline and three-month follow-up; baseline and six-month follow-up; discharge and three month-follow-up and three-month and six-month follow-up.
2. To assess the internal and external responsiveness of the FIM between baseline and discharge; baseline and three-month follow-up; baseline and six-month follow-up; discharge and three month-follow-up and three-month and six-month follow-up.

4.5 Method

4.5.1 Study design

This study was a clinical measurement study completed as part of a pilot RCT which examined the effectiveness of mirror therapy in the sub-acute stage of

stroke. The psychometric properties of the outcome measures assessed were internal and external responsiveness.

4.5.2 Participants

Participants were recruited across three stroke rehabilitation units, as part of a pilot RCT. The inclusion and exclusion criteria are those reported in Chapter 3. The full study sample of 40 participants were included in the analysis, with recruitment taking place between May 2015 and December 2017. Following randomisation, participants received either mirror therapy alongside standard occupational therapy treatment or standard occupational therapy treatment, during their inpatient rehabilitation.

4.5.3 Ethical approval

Ethical approval was granted by the Office for Research and Ethics Committees, as reported in Chapter 3. Written informed consent was gained from participants, with additional written informed consent gained for video recording of the gWMFT.

4.5.4 Outcome measures

4.5.4.1 Functional Independence Measure

The FIM is an 18-item clinician-reported outcome measure, measuring performance in ADLs and aspects of cognition (Uniform Data System for Medical Rehabilitation 1997). The FIM aspect of the UK FIM+FAM version 2.2 (UK FIM+FAM Users Group 2011) was used in the current study. The motor and cognitive sub-scales consist of 13 and five items respectively; areas of self-care, sphincter control, mobility, locomotion, communication and social

cognition are assessed. Using a seven-point ordinal scale, participants are scored between one (full assistance) and seven (full independence) on each item, indicating the level of assistance required. The sum of each sub-scale reflects a summary score for each domain and added together these represent a total FIM score.

4.5.4.2 graded Wolf Motor Function Test

The gWMFT is an assessment of upper limb motor function, encompassing joint-specific and increasingly complex upper limb movements (Constraint Induced Movement Therapy Research Group 2002). The gWMFT comprises 13 items, each with a level A and level B, enabling task adjustment depending on participant ability (Constraint Induced Movement Therapy Research Group 2002) (Appendix 3). Level A is more demanding of participant ability than Level B. The gWMFT is described in detail in Chapters 2 and 3.

4.5.5 Scoring and procedure

Scoring was completed by a research occupational therapist. Formal training on delivery and scoring of the assessments was completed prior to commencement of the study. The same researcher assessed all participants, blinded to allocation of treatment group. Assessments at three- and six-month follow-up were generally completed at participants' homes. Baseline measurements took place within one week of recruitment, followed by assessments at discharge, and at three- and six-months post-baseline (see Appendix 10 for the flow diagram of study procedure).

For scoring of the FIM, participants were encouraged to complete all personal ADLs to the best of their ability. Assessments at baseline and discharge were

completed at the participant's hospital bedside. Scoring was based on those activities observed by the researcher, with consideration for use of equipment, time taken to complete tasks and the level of assistance required.

The gWMFT was administered according to protocol guidelines (Constraint Induced Movement Therapy Research Group 2002). Assessments at baseline and discharge took place in a quiet room used for research purposes in the hospital. This enabled standardised placement of the equipment and video camera. Placement of objects and participants were standardised through the use of a plexiglass template, which was devised according to protocol instructions and securely affixed to a table-top (Constraint Induced Movement Therapy Research Group 2002). Participants were seated at the table in a wheelchair and their upper limb activities were videotaped. The research occupational therapist scored participants through direct observation.

4.5.6 External criterion

Upon discharge participants completed an exit questionnaire, which formed the criterion for external responsiveness. Participants were asked, "In your opinion do you feel that your occupational therapy sessions have been of any benefit with regard to the return of movement to your affected arm?", and optional answers were 'very beneficial', 'quite beneficial', 'of little benefit' or 'of no benefit'. Responses were numbered, one to four, with one representing 'no benefit' to four representing 'very beneficial'. This was transformed into a dichotomous scale, to denote patient-reported rating of change, for assessing external responsiveness.

4.5.7 Data analysis

Participant results were collated into an Excel spreadsheet. Following this the data was cleaned and entered into SPSS for further analysis. It is expected that following stroke individuals experience some degree of spontaneous recovery, with the aim of stroke rehabilitation to supplement and/or augment this process (Cramer 2008). Therefore, it was assumed that participants would experience improvements in function over the study period. The pilot RCT was not powered to detect statistical significance and the aim of this study was to investigate the responsiveness of the outcome measures and not treatment efficacy. However, due to participant randomisation, results are reported for the total sample and treatment groups for reference.

Descriptive statistics were reported for clinical characteristics of gender, age and side of hemiplegia. Linear mixed models, including baseline scores as covariates, were completed to adjust for missing values at random for both outcome measures. Estimated marginal means (EMM) and standard error (SE) values were reported for outcome measures completed at discharge, and at three- and six-month follow-up, for the total sample and according to treatment group.

4.5.7.1 Floor and ceiling effects

Floor and ceiling effects were examined using score distributions at baseline, discharge and both follow-up periods. Floor effects were indicated when 15% or more of the sample achieved the minimum or maximum of scores (McHorney and Tarlov 1995).

4.5.7.2 Internal responsiveness

Internal responsiveness was assessed using ES and SRM, effect size statistics. The ES was calculated by dividing the mean difference in scores by the standard deviation of the first assessment score for the time interval under investigation (Husted *et al*. 2000). Effect size values less than 0.2 were considered small, values between 0.5 and 0.8 were considered moderate and values above 0.8 were considered large (Norman *et al*. 2007). The SRM is computed by dividing the mean difference in scores by the standard deviation of the mean change (Liang *et al*. 1990; Husted *et al*. 2000). These were completed for the total sample and for the intervention and control groups for each time interval; between baseline and discharge, discharge and three-month assessments and between three- and six-month assessments. Higher values of ES and SRM were indicative of a greater degree of responsiveness. Positive values were seen as improvements in ability for the FIM and gWMFT functional ability scores, and negative values for gWMFT performance time were indicative of an improvement in performance.

4.5.7.3 External responsiveness

To dichotomise the patient-reported rating of change, those who reported treatment as “very beneficial” were classified under important change and the remaining responses under unimportant change. Page *et al*. (2012) used a similar method to dichotomise responses on a global rating scale, where only those who reported “excellent improvement” were classified under meaningful improvement.

The association between the external criterion and change in outcome measure scores at baseline and discharge were initially examined using Spearman rank

correlation coefficients. Correlations greater than 0.3 were considered an adequate association (Revicki *et al*. 2008).

Receiver operating characteristic (ROC) curves were completed where the change in score correlated adequately with the external criterion. The ROC method is often used to denote the ability of an outcome measure to discriminate between patients who have improved and those who have not improved (Deyo and Centor 1986; Husted *et al*. 2000).

The ROC curve is a plot of sensitivity identifying true change, versus 1 minus specificity, identifying false positives (Hanley and McNeil 1982). The area under the curve (AUC) represents the probability of the outcome measure to correctly identify improvement. Values at 1 represent perfect ability of the outcome measure to differentiate improvement, and values at 0.5 and less indicates no ability to discriminate (Deyo and Centor 1986; Angst *et al*. 2008). An AUC greater than 0.7 was considered an adequate indicator of external responsiveness (Terwee *et al*. 2007).

Data analysis was completed using SPSS Version 25.0 (SPSS Inc. Chicago) and Microsoft Excel 2016.

4.6 Results

Overall, forty participants were recruited to the pilot RCT and randomised equally. Figure 4.1 presents the participant flow through the pilot RCT. The demographic characteristics of the participants at baseline are reported in Table 4.1. The number of males in either treatment group almost doubled that of the

number of females recruited. Participants were approximately 15 days post-stroke and predominantly experienced left sided hemiplegia. The gWMFT was not completed at three- and six-month follow-up for two participants, due to difficulty setting up equipment for this outcome measure in their homes.

Figure 4.1 Flow of participants through the study

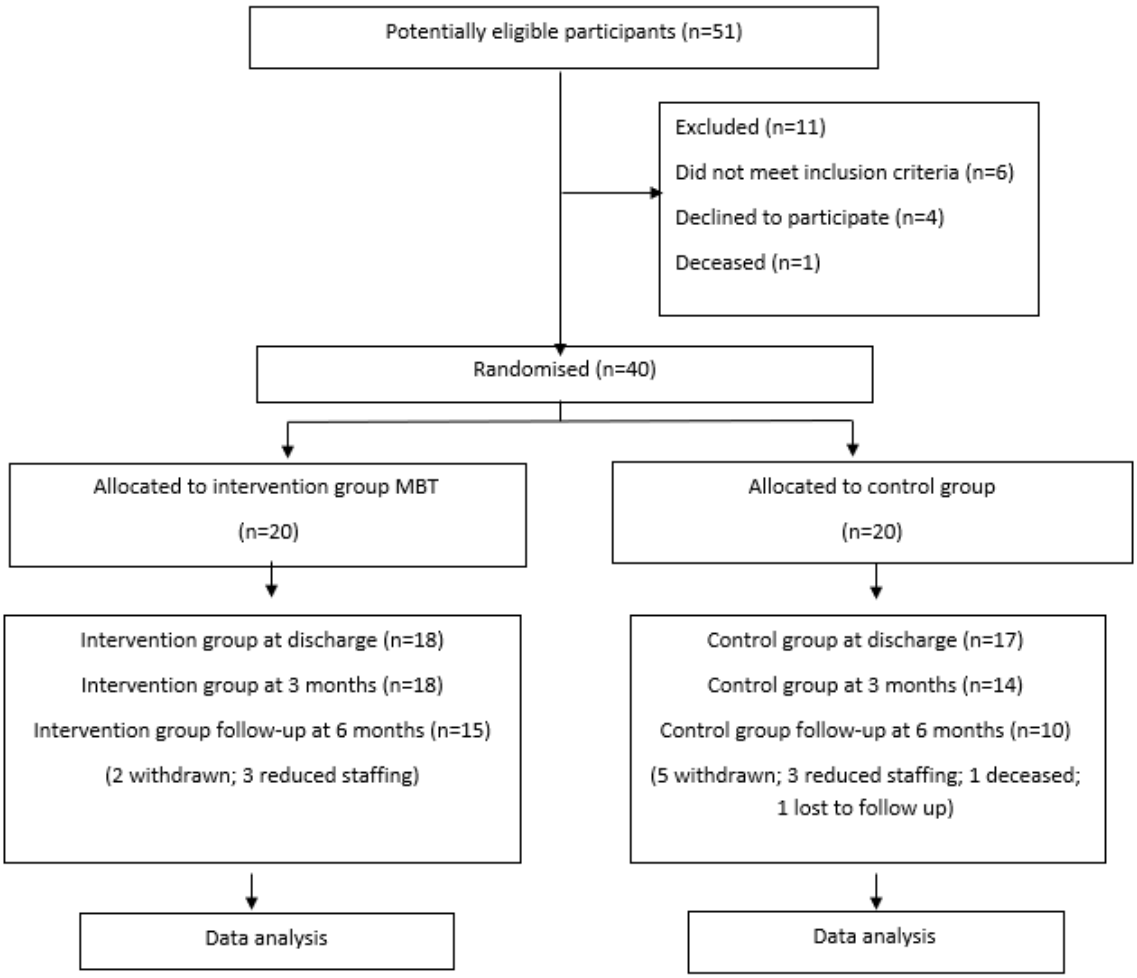


Table 4.1 Baseline characteristics of participants

	Total (n=40)	**Intervention (n=20)**	**Control (n=20)**
Age in years, mean (SD)	71.25 (9.23)	72.75 (9.55)	69.75 (8.9)
Days post-stroke, mean (SD)	15.6 (8.82)	16.5 (9.38)	14.7 (8.36))
Gender			
Male, n (%)	27 (67.5%)	14 (70%)	13 (65%)
Female, n (%)	13 (32.5%)	7 (35%)	7 (35%)
Side of hemiplegia			
Left, n (%)	24 (60%)	10 (50%)	14 (70%)
Right, n (%)	16 (40%)	10 (50%)	6 (30%)
FIM total			
Mean (SD)	64.35 (9.28)	63.85 (11.34)	64.85 (6.91)
Floor effect, n (%)	0	0	0
Ceiling effect, n (%)	0	0	0
FIM motor			
Mean (SD)	31.4 (7.11)	31.85 (7.57)	30.95 (6.80)
Floor effect, n (%)	0	0	0
Ceiling effect, n (%)	0	0	0
FIM cognitive			
Mean (SD)	32.95 (3.73)	32.00 (5.05)	33.90 (1.07)
Floor effect, n (%)	0	0	0
Ceiling effect, n (%)	10 (25)	5 (25)	5 (25)
gWMFT FAS			
Mean (SD)	2.73 (2.27)	2.63 (2.22)	2.83 (2.38)
Floor effect, n (%)	16 (40)	8 (40)	8 (40)
Ceiling effect, n (%)	0	0	0
gWMFT time			
Mean (SD)	58.92 (56.00)	66.53 (54.74)	51.32 (57.61)
Floor effect, n (%)	17 (42.5)	9 (45)	8 (40)
Ceiling effect, n (%)	0	0	0

Abbreviations: FIM: Functional Independence Measure; gWMFT: graded Wolf Motor Function Test; FAS: Functional Ability Scale.

The EMMs according to time and treatment group, controlling for baseline values, are reported in Table 4.2 for the FIM total, FIM motor sub-scale and FIM cognitive sub-scale and Table 4.3 for the gWMFT functional ability scores and gWMFT performance time.

Table 4.2 Estimated marginal means (EMM) and standard error (SE) at discharge, three-month and six-month follow-up for the Functional Independence Measure

Time point		EMM (SE)			Floor effect n (%)			Ceiling effect n (%)		
		FIM total	FIM motor	FIM cognitive	FIM total	FIM motor	FIM cognitive	FIM total	FIM motor	FIM cognitive
Discharge	**Full Sample (n=35)**	79.07 (2.22)	45.24 (2.15)	33.85 (0.21)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	8 (22.9)
	Intervention (n=18)	75.6 (3.14)	41.37 (3.05)	33.83 (0.29)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (16.7)
	Control (n=17)	82.54 (3.20)	49.11 (3.11)	33.87 (0.30)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5 (29.4)
Three-month	**Full sample (n=32)**	96.91 (2.31)	63.24 (2.23)	33.70 (0.22)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	6 (18.8)
	Intervention (n=18)	94.71 (3.14)	60.54 (3.05)	33.78 (0.29)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (11.1)
	Control (n=14)	99.10 (3.41)	65.94 (3.31)	33.63 (0.33)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (28.6)
Six-month	**Full sample (n=25)**	99.78 (2.52)	66.23 (2.44)	33.59 (0.25)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5 (20)
	Intervention (n=15)	96.54 (3.31)	62.68 (3.21)	33.46 (0.32)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (20)
	Control (n=10)	103.01 (3.84)	69.78 (3.74)	33.71 (0.38)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (20)

Abbreviations: EMM, estimated marginal mean; SE, standard error; %, percentage; FIM, Functional Independence Measure

Table 4.3 Estimated marginal means (EMM) and standard error (SE) at discharge, three-month and six-month follow-up for the graded Wolf Motor Function Test

Time point		EMM (SE)		Floor effect n (%)		Ceiling effect n (%)	
		gWMFT functional ability	gWMFT performance time	gWMFT functional ability	gWMFT performance time	gWMFT functional ability	gWMFT performance time
Discharge	**Full Sample (n=35)**	4.02 (0.25)	43.67 (5.59)	5 (14.3)	11 (27.5)	1 (2.9)	0 (0)
	Intervention (n=18)	3.88 (0.35)	45.32 (7.85)	2 (11.1)	6 (33.3)	2 (11.1)	0 (0)
	Control (n=17)	4.16 (0.36)	42.01 (7.95)	3 (17.6)	5 (29.4)	1 (5.9)	0 (0)
Three-month	**Full sample (n=32)**	4.72 (0.26)	35.49 (5.82)	3 (13.3)	7 (17.5)	6 (20)	0 (0)
	Intervention (n=18)	4.3 (0.35)	41.25 (7.85)	2 (11.1)	5 (27.8)	2 (11.1)	0 (0)
	Control (n=12)	5.14 (0.38)	29.73 (8.59)	1 (8.3)	2 (16.7)	4 (33.3)	0 (0)
Six-month	**Full sample (n=23)**	4.98 (0.27)	32.95 (6.12)	1 (4.3)	6 (15)	8 (34.8)	0 (0)
	Intervention (n=14)	4.59 (0.37)	36.75 (8.19)	0 (0)	4 (28.6)	4 (28.6)	0 (0)
	Control (n=9)	5.37 (0.40)	29.16 (9.10)	1 (11.1)	2 (22.2)	4 (44.4)	0 (0)

Abbreviations: EMM, estimated marginal mean; SE, standard error; %, percentage; gWMFT, graded Wolf Motor Function Test

4.6.1 Floor and ceiling effects

Ceiling effects were found across most assessment periods for FIM cognitive scores for the total sample and treatment groups (see Tables 4.1 and 4.2). Floor effects were found at baseline for the gWMFT FAS for the total sample, with 40% of participants achieving a score less than one (Table 4.1). Substantial ceiling effects were found for the gWMFT FAS at three-month follow-up for the total sample and control group, and for the total sample and both treatment groups at six-month follow-up (Table 4.3). Floor effects were demonstrated across all assessment periods for gWMFT performance time for the total sample and both treatment groups.

4.6.2 Functional Independence Measure

4.6.2.1 Internal responsiveness

The ES and SRM values for the total sample, intervention group and control group are reported in Table 4.4. Positive responsiveness values were found for the FIM total and FIM motor sub-scale across all time intervals, indicating an improvement in function. Large ES were found for the FIM total and FIM motor sub-scale between baseline and discharge, and between discharge and three-month follow-up, with small values found by six-month follow-up (Table 4.4). The corresponding SRM values were similar up to three-month follow-up. with moderate values found between three- and six-month follow-up. Responsiveness of the FIM total and FIM motor scores remained large for the control group for all time intervals (Table 4.4). Responsiveness analyses for the FIM cognitive sub-scale demonstrated negligible to minor effect sizes for the total sample across all time intervals, with negative values found after discharge, potentially indicating a worsening in ability.

Table 4.4 Mean difference and responsiveness values across each time interval for the Functional Independence Measure

Time point		Mean difference (SD)			ES			SRM		
		FIM total	FIM motor	FIM cognitive	FIM total	FIM motor	FIM cognitive	FIM total	FIM motor	FIM cognitive
Baseline to discharge	Full Sample (n=35)	13.6 (8.74)	13.46 (8.65)	0.14 (0.88)	1.94	2.07	0.1	1.56	1.56	0.16
	Intervention (n=18)	10.34 (7.45)	10.67 (7.12)	0.28 (1.23)	0.91	1.64	0.15	1.39	1.50	0.23
	Control (n=17)	16.41 (9.33)	16.41 (9.33)	0 (0)	2.59	2.69	0	1.76	1.76	0
Discharge to three-month	Full sample (n=31)	17.81 (14.83)	17.97 (14.65)	-0.16 (0.78)	1.53	1.61	-0.19	1.20	1.23	-0.21
	Intervention (n=18)	19.12 (12.49)	19.18 (12.58)	-0.06 (0.43)	1.77	1.8	-0.06	1.53	1.52	-0.14
	Control (n=13)	23.36 (32.68)	21.21 (25.16)	-0.31 (1.11)	1.34	1.36	-0.38	0.71	0.84	-0.28
Three-month to six-month	Full sample (n=25)	2.88 (4.87)	3 (4.41)	-0.12 (1.81)	0.15	0.16	-0.09	0.59	0.68	-0.07
	Intervention (n=15)	2.33 (5.69)	2.67 (5.29)	-0.33 (2.19)	0.13	0.14	-0.37	0.41	0.50	-0.15
	Control (n=10)	3.70 (3.40)	3.50 (2.8)	0.20 (1.03)	0.17	0.17	-0.1	1.09	1.25	0.19
Baseline to three-month	Full sample (n=32)	29 (21.69)	30.24 (18.19)	0.00 (1.21)	4.68	4.89	0	1.35	1.66	0
	Intervention (n=18)	30.06 (13.72)	29.83 (13.71)	0.22 (1.26)	2.65	4.59	0.12	2.19	2.18	0.18
	Control (n=14)	34.29 (18.71)	34.57 (18.17)	-0.31 (1.11)	5.18	5.29	-0.37	1.83	1.90	-0.28

Baseline to six-month	**Full sample (n=25)**	34.12 (17.88)	34.20 (17.18)	-0.08 (2.47)	4.91	5.13	-0.05	1.91	1.99	-0.03
	Intervention (n=15)	30.6 (16.25)	30.6 (15.08)	0 (3.21)	2.7	4.66	0	1.88	2.03	0
	Control (n=10)	39.40 (19.76)	39.60 (19.48)	-0.2 (0.42)	6.42	6.59	-0.24	1.99	2.03	-0.47

Abbreviations: SD, standard deviation; ES, effect size; SRM: standardised response mean; FIM: Functional Independence Measure

4.6.3 graded Wolf Motor Function Test

4.6.3.1 Internal responsiveness

The ES and SRM values for the total sample, intervention group and control group are reported in Table 4.5. Positive responsiveness values found for gWMFT functional ability scores and negative values found for the gWMFT performance time indicated an improvement in upper limb function across all time intervals. Standardised response mean values for the gWMFT functional ability scores were large between baseline and discharge and moderate between discharge and three-month follow-up for the total sample, with moderate and small ES values found for the corresponding time intervals (Table 4.5). An increase in SRM for gWMFT functional ability scores was seen between three- and six-month follow-up for the total sample, however corresponding ES values were negligible (ES <0.1) (Table 4.5). The responsiveness of the gWMFT performance time was small for the total sample between baseline and discharge for the time intervals up to three months, with reduction to a negligible SRM value at six months for the intervention group (Table 4.5).

Table 4.5 Mean difference and responsiveness values across each time interval for the graded Wolf Motor Function Test

Time point		Mean difference (SD)		ES		SRM	
		gWMFT FAS	gWMFT time	gWMFT FAS	gWMFT time	gWMFT FAS	gWMFT time
Baseline to discharge	**Full Sample (n=35)**	1.31 (1.27)	-15.95 (33.76)	0.57	-0.29	1.03	-0.47
	Intervention (n=18)	1.13 (1.15)	-14.52 (32.12)	0.52	-0.3	0.98	-0.45
	Control (n=17)	1.50 (1.40)	-17.62 (37.32)	0.61	-0.3	1.07	-0.31
Discharge to three-month	**Full sample (n=29)**	0.62 (1.11)	-6.81 (24.16)	0.26	-0.12	0.56	-0.28
	Intervention (n=18)	0.41 (0.80)	-4.07 (13.59)	0.16	-0.07	0.52	-0.30
	Control (n=11)	1.38 (2.01)	-10.51 (34.38)	0.40	-0.21	0.68	-0.30
Three-month to six-month	**Full sample (n=23)**	0.24 (0.33)	-2.48 (11.15)	0.07	-0.05	0.72	-0.22
	Intervention (n=14)	0.24 (0.30)	-3.99 (14.29)	0.08	-0.07	0.81	-0.07
	Control (n=9)	0.24 (0.40)	-1.26 (3.07)	0.09	-0.0	0.6	-0.41
Baseline to three-month	**Full sample (n=30)**	1.96 (1.79)	-23.67 (40.68)	0.87	-0.43	1.09	-0.58
	Intervention (n=18)	1.54 (1.46)	-18.59 (32.84)	0.71	-0.34	1.06	-0.57
	Control (n=12)	2.58 (2.12)	-31.88 (50.60)	1.07	-0.54	1.22	-0.63

Baseline to six-month	**Full sample (n=23)**	2.26 (2.04)	-25.54 (42.62)	0.95	-0.43	1.11	-0.60
	Intervention (n=14)	1.88 (1.64)	-23.45 (38.12)	0.72	-0.42	1.15	-0.62
	Control (n=9)	2.85 (2.53)	-28.80 (51.12)	1.12	-0.48	1.13	-0.56

Abbreviations: SD, standard deviation; ES, effect size; SRM: standardised response mean; gWMFT: graded Wolf Motor Function Test; FAS, functional ability scale.

4.6.3.2 External responsiveness

The exit questionnaire was completed upon discharge from hospital by 32 participants, with 68.75% (n=23) of respondents classified under important change. The remaining participants (n=9) reported occupational therapy treatment was ‘quite beneficial’ and were classified under no important change. The mean scores for participants categorised according to the external criterion are shown in Table 4.6. The important change group demonstrated greater improvements than the unimportant change group for the FIM total, FIM motor, gWMFT functional ability scores and gWMFT performance time. There were minimal differences between groups for FIM cognitive scores.

Table 4.6 Mean scores and change scores, according to the external criterion, and correlations between the external criterion and change score

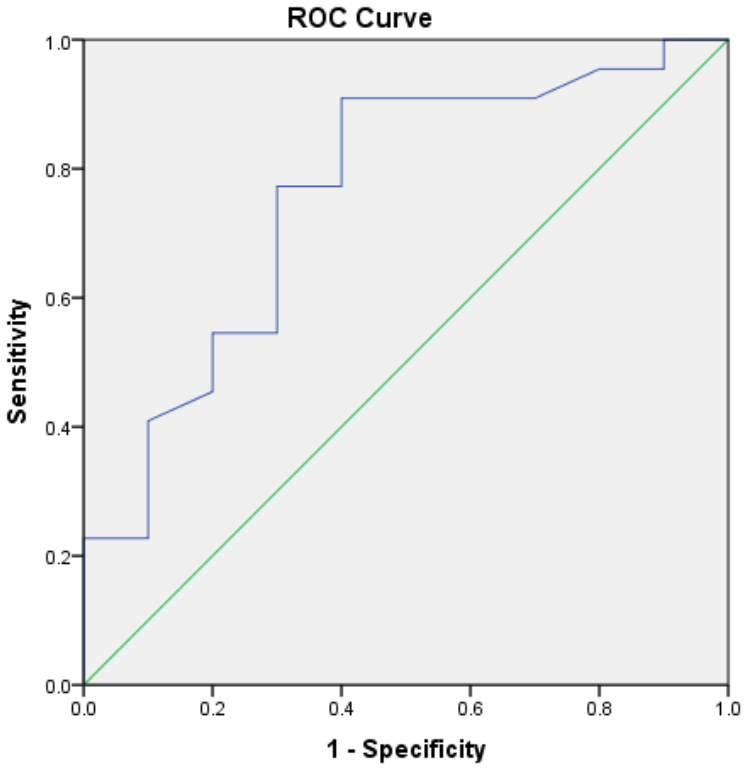
	Baseline Mean (SD)		**Discharge Mean (SD)**		**Mean change score (SD)**		**External criterion Spearman's rho (95% CI)**
	Important (n=22)	**Not important (n=10)**	**Important (n=22)**	**Not important (n=10)**	**Important (n=22)**	**Not important (n=10)**	**(n=32)**
FIM total	65.68 (7.47)	62.00 (4.67)	79.32 (11.03)	74.10 (10.44)	13.64 (8.52)	12.10 (9.55)	0.128 (-0.246-0.465)
FIM motor	32.27 (6.71)	27.80 (4.24)	45.59 (10.87)	40.10 (10.01)	13.32 (8.47)	12.3 (9.36)	0.095 (-0.271-0.454)
FIM cognitive	33.41 (1.68)	34.20 (1.23)	33.73 (0.88)	34.00 (1.25)	0.32 (0.99)	-0.2 (0.63)	0.263 (-0.022-0.474)
gWMFT FAS	2.90 (2.34)	1.61 (1.86)	4.52 (2.21)	2.21 (2.35)	1.62 (1.42)	0.6 (0.75)	0.416[b] (0.093-0.454)
gWMFT time	56.88 (56.28)	83.21 (52.89)	33.29 (50.42)	79.42 (55.83)	-23.6 (41.46)	-3.79 (6.52)	-0.171 (-0.499-0.172)

Abbreviations: SD, standard deviation; 95% CI, 95% confidence interval; FIM, Functional Independence Measure; gWMFT, graded Wolf Motor Function Test; FAS Functional Ability Scale.

[b] Significant at 0.05 level (2-tailed)

The gWMFT functional ability scale was the only outcome measure which correlated significantly with the external criterion (Spearman's Rho = 0.416, p=0.018) (Table 4.6). A ROC curve was produced for the gWMFT FAS (Figure 4.2), with an AUC value of 0.740 (95% confidence interval, 0.541-0.938).

Figure 4.2 Receiver operating characteristic curve for change in the graded Wolf Motor Function Test Functional Ability Scale scores for participants classified under important change (area under the curve = 0.74)



4.7 Discussion

This study examined the responsiveness of the FIM and gWMFT as part of a pilot RCT using both internal and external responsiveness methods. The FIM total, FIM motor and gWMFT, demonstrated adequate internal responsiveness, with negligible responsiveness found for FIM cognitive scores. The gWFMT FAS was the only outcome measure to demonstrate external responsiveness using patient-reported rating of change. The results of this study were used to inform the development of a

large-scale RCT investigating the effectiveness of mirror therapy among patients in the sub-acute phase of stroke.

4.7.1 Functional Independence Measure

There were no floor or ceiling effects found for the FIM total and FIM motor sub-scale. Participants in the current study demonstrated lower levels of functional ability using the FIM total and FIM motor sub-scale in comparison to the FIM responsiveness studies completed by Van Der Putten *et al*. (1999) and Schepers *et al*. (2006). This could be the result of participants recruited and assessed in the earlier stages of stroke, with the length of time between date of stroke and first assessment ranging from four days to 43 days. In comparison participants ranged from 15 to 129 days post-stroke in the study by Schepers *et al*. (2006), with no information provided regarding time post-stroke in the study by Van Der Putten *et al*. (1999). Scores for the FIM cognitive sub-scale approached a ceiling effect in the study by Van Der Putten *et al*. (1999). However, Schepers *et al*. (2006) reported ceiling effects for the FIM cognitive sub-scale at baseline, six months and 12 months post-stroke. Overall, these studies highlighted issues with the use of the FIM cognitive sub-scale in a stroke population.

In the current study the eligibility criteria may have contributed to the lack of variability demonstrated in cognitive scores. Participants were recruited as part of an upper limb intervention study and were required to demonstrate the cognitive ability to engage in mirror therapy. In addition, stroke-related cognitive impairment may not be adequately assessed by the FIM cognitive scale. Cognitive impairment is commonly experienced after stroke (Jin *et al*. 2006) and the use of a more robust cognitive screening tool in a main RCT may be warranted.

Internal responsiveness

The FIM total and FIM motor sub-scale demonstrated large levels of responsiveness using ES and SRM up to three months follow-up, with the highest level of responsiveness found between baseline and discharge. The findings are similar to studies completed with stroke survivors during inpatient rehabilitation and supports the use of the FIM in a main RCT and in stroke rehabilitation (Van Der Putten *et al*. 1999; Wallace *et al*. 2002; Schepers *et al*. 2006). Van Der Putten *et al*. (1999) assessed the responsiveness of the FIM with 82 stroke survivors using the standardised ES. For the duration of inpatient rehabilitation Van Der Putten *et al*. (1999) reported ES values of 0.82 for the FIM total, 0.91 for the FIM motor sub-scale and 0.61 for the cognitive sub-scale. Schepers *et al*. (2006) assessed the responsiveness of the FIM up to 12 months post-stroke in 308 participants and reported ES of 0.84 and 0.89 for the FIM total and FIM motor sub-scale between baseline and six months post-stroke. The higher levels of responsiveness found in the current study for the FIM total and FIM motor sub-scale may be the result of the lower FIM scores found at baseline, with participants demonstrating a higher degree of functional impairment. Passalent *et al*. (2011) reported admission FIM scores were significant predictors of scores at discharge. Participants categorised with low FIM scores following admission to rehabilitation were more likely to demonstrate the greatest amount of change (Passalent *et al*. 2011).

Time since stroke and time between assessment periods are also likely contributory factors to the differences in responsiveness found (Wallace *et al*. 2002). In the current study participants were approximately 15 days post-stroke when baseline assessments were completed, and the mean length of time spent in inpatient

rehabilitation was 38.8 days (SD=23.24). In the study by Van Der Putten *et al*. (1999) the mean length of time spent in hospital was 49.5 days (SD=35.1). Participants were a median of 41 days post stroke when first assessed, and the mean length of time spent in hospital was 81 days (SD=33) in the study by Schepers *et al*. (2006). The initial time interval assessed by Schepers *et al*. (2006) was between baseline and six months post-stroke, with larger values of responsiveness found in the current study for the same time period.

The variability in ES and SRM values demonstrated between groups and by time, may have been influenced by the reduction in sample size across each testing interval, with 25 participants remaining by six-month follow-up. Reduced effect sizes for FIM total and FIM motor scores have previously been found as time increased post-stroke (Wallace *et al*. 2002; Schepers *et al*. 2006). Furthermore, the variations in functional ability is often indicative of the heterogeneous nature of stroke recovery which generally plateaus at six months' following onset, potentially leading to the reductions in ability to detect change over time (Kwakkel and Kollen 2013).

Due to ceiling effects the FIM cognitive sub-scale was potentially unable to capture any change which took place across the study thus leading to the negligible levels of responsiveness found at all time intervals. In previous studies lower levels of responsiveness have been reported for the FIM cognitive sub-scale in comparison to the FIM total and FIM motor sub-scale (Van der Putten *et al*. 1999; Hobart *et al*. 2001; Schepers *et al*. 2006). Although ceiling effects were reported for the FIM cognitive sub-scale by Schepers *et al*. (2006), moderate levels of responsiveness were reported, with an ES of 0.47 found at all time intervals. Hobart *et al*. (2001) reported an SRM value of 0.17 for the FIM cognitive sub-scale following inpatient rehabilitation, similar

to the level of responsiveness found in the current study. However, participants with stroke and multiple sclerosis were included in the analysis with participants at variable time points post-stroke which limits comparisons.

External responsiveness

Changes in the FIM total, FIM motor sub-scale and FIM cognitive sub-scale between baseline and discharge did not correspond with the external criterion. The external criterion used was specifically linked to improvement in participant arm function. As such, changes in overall ADL function may not be associated with patient-perceived changes in arm function. The use of an external criterion that reflects the domains captured by the FIM would be advisable for future research.

The MCID can also be examined as an indicator of external responsiveness (Crosby *et al*. 2003). A study by Wallace *et al*. (2002) used the Modified Rankin Scale as an external criterion, where a change of one level or more represented important change on the FIM motor scale between one-month and three-months post-stroke. This study found a difference of 11 points on the FIM motor sub-scale to be representative of the MCID, which is the smallest change in scores considered meaningful, with an AUC of 0.675 (Wallace *et al*. 2002). Beninato *et al*. (2006) examined the MCID of the FIM in 113 stroke survivors who were discharged from stroke rehabilitation using perceptions of change in function by the participants doctor. Scores changes of 22, 17 and 3 found for the FIM total, FIM motor sub-scale and FIM cognitive sub-scale respectively were considered indicative of the MCID (Beninato *et al*. 2006). Due to the different external criterions used, MCID values reported for the FIM motor sub-scale were different across studies (Beninato *et al*. 2006; Wallace *et al*. 2002), with a greater degree of change required to be considered meaningful in the study by Beninato *et al*. (2006).

Although the patient-reported external criterion was not applicable for the FIM, the current study supports findings that variations in external responsiveness generally occurs depending on the type of criterion used (Beaton 2000). The determination of clinically important change will vary according to the criterion used, time post-stroke, duration between assessments, stage of rehabilitation and baseline score (Beaton 2000; Husted *et al*. 2000; Wallace *et al*. 2002; Beninato *et al*. 2006).

4.7.2 graded Wolf Motor Function Test

Scores for the gWMFT performance time demonstrated substantial floor effects across all testing intervals. The gWMFT functional ability scores demonstrated floor effects at baseline and ceiling effects at six-month follow-up for both treatment groups. The ceiling effect for functional ability by six months indicated the limited utility of the gWMFT with individuals in the chronic phase of stroke. The floor and ceiling effects for the gWMFT were also noted in Chapter 3. There were no standardised restrictions on the upper limb ability of participants included in the current study, which may reflect the floor effects found at baseline for the gWMFT. Discussion of the floor and ceiling effects for the gWMFT in Chapter 3 remain applicable following analysis using the total sample in the current Chapter.

Internal responsiveness

This is the first study to examine the responsiveness of the gWMFT. The gWMFT was most responsive between baseline and discharge, with SRM values of 1.03 found for functional ability and -0.47 found for performance time for the total sample. The SRM values for functional ability continued to indicate improvement in upper limb function up to six months with an SRM of 0.56 between discharge and three months and an SRM of 0.72 by six months. However, moderate values of ES were found for

functional ability between baseline and discharge for the total sample (ES=0.57), with small ES found for performance time (ES=-0.29). Low levels of responsiveness were found for functional ability at subsequent time intervals with an ES of 0.26 reported by three-months and an ES of 0.07 found by six months. Responsiveness scores also declined for performance time as the study progressed. The small sample size and the recruitment of participants with varying levels of upper limb ability may have increased variability in scores on the gWMFT thereby reducing ES values.

Although, comparison with other studies is limited, the responsiveness of the WMFT has been assessed in the early stages of stroke. Edwards *et al*. (2012) examined the responsiveness of the WMFT with 51 participants recruited within 28 days of stroke onset. Effect sizes greater than one were found for WMFT functional ability scores between baseline and 14 days later, and between baseline and 90 days post-stroke. Contrary to the current study, the responsiveness of performance time was moderate (ES=0.61) between baseline and 14 days, and large (ES=0.85) between baseline and 90 days post-stroke (Edwards *et al*. 2012). In addition, the study by Edwards *et al*. (2012) was completed as part of a trial examining the effectiveness of constraint-induced movement therapy and participants were required to demonstrate a minimum level of function. The inclusion criteria for participants may have led to reduced variability in scores leading to higher levels of responsiveness (Edwards *et al*. 2012). Lin *et al*. (2009a) investigated the responsiveness of WMFT functional ability scores using ES at several time points post-stroke, ranging from 14 days to 180 days post-stroke, with similar levels of responsiveness found.

Consistent with previous psychometric evaluation of the WMFT (Lin *et al*. 2009b; Edwards *et al*. 2012), higher responsiveness was found for the gWMFT functional

ability scores compared to performance time. The floor effects found for performance time scores potentially reduced its responsiveness, due to its limited ability to capture the full scope of participant upper limb function. However, comparison between studies should be approached with caution due to different versions of the WMFT delivered, the different time intervals applied between assessment periods and different inclusion criteria for participants.

External responsiveness

The gWMFT functional ability scale was the only domain sensitive to patient-perceived change, with an adequate AUC of 0.74. This indicates that change on the FAS of the gWMFT between baseline and discharge reflected meaningful change in upper limb function as perceived by patients.

Patient-reported rating of change was used to estimate the MCID of the WMFT in the early stages of stroke (days post stroke, mean=9.5) (Lang *et al*. 2008). Participants were assessed at baseline and 14 days later and most participants reported their upper as much better (Lang *et al*. 2008), consistent with findings in the current study. If the stroke affected the participant's dominant side the MCID was 19 seconds for performance time and 1.0 for functional ability (1.2, if non-dominant side affected) (Lang *et al*. 2008). Conversely, Lin *et al*. (2009b) investigated the MCID of the WMFT with participants six months post-stroke before treatment and three weeks later. A change of between six to ten points on the Fugl-Meyer assessment was indicative of clinically relevant change on the WMFT and the reported MCID was 1.37 for performance time and 0.33 for functional ability (Lin *et al*. 2009b). These differences are likely the result of the different criterions implemented and the different stages of recovery at which the participants were at (Lang *et al*. 2008; Lin *et al*. 2009b). Although

improvement in recovery is generally more pronounced in the earlier stages of stroke, greater levels of change were required on the FAS for participants to consider this improvement as meaningful.

As noted earlier in the discussion section of this Chapter, methodological variations in how responsiveness is defined, method of assessment, the external criterion reported and the sample assessed, limits the ability to compare responsiveness across studies (Husted *et al*. 2000; Terwee *et al*. 2003). In addition, the ROC method has not been used to investigate the responsiveness of the gWMFT functional ability. There are variations in the type of external criterion used across studies, all of which are context specific (Beninato *et al*. 2006; Lang *et al*. 2008; Lin *et al*. 2009b). However, patient-perceived criterions are often viewed as the main standard by which meaningful change can be gauged, due to the direct relevance to the patient (Revicki *et al*. 2006; Lang *et al*. 2008).

4.7.3 Effect size and standardised response mean

Larger SRM values were generally found across all time intervals in comparison to the standardised ES. The responsiveness values derived from these methods use different denominators and the varying results found for ES and SRM are to be expected (Norman *et al*. 2007). The lower levels of ES indicated there was a high degree of variability in participant scores for the initial assessment for each time interval. Although guidelines for the interpretation of ES values have been used to interpret SRM values, this study highlights the difficulty this poses due to the varying levels of responsiveness found according to the method of analysis used (Middel and Van Sonderen 2002). Husted *et al*. (2000) and Guyatt *et al*. (1987) reported a preference for SRM over standardised ES because the denominator is a measure of

the variability of change in participant scores. However, there remains no agreed method for evaluating responsiveness (Norman *et al*. 2007).

4.7.4 Limitations

Participants were recruited as part of a pilot RCT, which examined a small sample size of patients in the sub-acute stage of stroke, limiting study generalisability. By six-month follow-up approximately 50% and 75% of the sample initially recruited to the control and intervention groups respectively remained in the trial. This may have impacted on results, with participants with a greater degree of impairment leaving the study

The potential to recruit participants with more severe cognitive impairment was reduced due to inclusion criteria, which may have contributed to the lack of variability demonstrated across FIM cognitive scores. However, mirror therapy treatment requires cognitive ability to pay attention to the mirror image and instructions, with belief in the visual illusion potentially contributing to treatment efficacy (McCabe 2011). Recommendations for future study would be to include a cognitive screening tool, sensitive to the impact of vascular cognitive impairment.

The use of a retrospective patient-perceived criterion for external responsiveness has been criticised due to the potential for bias (Norman *et al* 1997). Recall bias occurs when the rating of change is made in relation to a person's current health status, with no connection to their baseline level of ability (Norman *et al*. 1997). In addition, most participants considered their occupational therapy treatment to be very beneficial, with few individuals reporting little or no change. Although participant responses were likely impacted by the significant degree of recovery experienced for the time period, the use of a 7- to 11- point Likert scale may allow for a more nuanced response (Kamper

et al. 2013). In addition, multiple anchors for the examination of external responsiveness, alongside the evaluation of internal responsiveness, are recommended in future study (Revicki *et al*. 2008).

4.8 Conclusion

The FIM and gWMFT demonstrated acceptable internal responsiveness for use with stroke survivors within three months of stroke onset with concerns highlighted for the use of the FIM cognitive scale. This study supports the use of the gWMFT as a measure of upper limb function in the sub-acute phase of stroke and highlighted the reduced clinical utility of the gWMFT in a community setting. The gWMFT FAS was found to be sensitive to patient-perceived change in upper limb function, providing additional insight for use in clinical practice. However, due to the presence of floor and ceiling effects for the gWMFT caution is encouraged in the interpretation of results.

Chapter 5

Chapter 5 - Responsiveness of the EQ-5D-5L and Canadian Occupational Performance Measure in a pilot randomised controlled trial

5.1 Abstract

Introduction

The EQ-5D-5L and Canadian Occupational Performance Measure (COPM) are patient-reported outcome measures (PROMs) and are considered to capture meaningful and important outcomes for stroke survivors and clinicians. The EQ-5D-5L assesses health-related quality of life (HRQOL) and the COPM captures client perceptions of occupational performance. This chapter aimed to determine the responsiveness of the EQ-5D-5L and COPM following inpatient stroke rehabilitation.

Method

Thirty-nine participants recruited as part of the pilot RCT were included in the analysis. The EQ-5D-5L was completed at baseline, discharge and at three- and six-month follow-up. EQ-5D-5L index values were computed using the crosswalk method (Van Hout *et al*. 2015) and participants rated their overall health using a visual analogue scale (VAS). The COPM was administered at baseline and at three- and six-month follow-up; participants prioritised occupational problems and rated their performance and satisfaction in these areas. Responsiveness was assessed using the standardised effect size and standardised response mean (SRM).

Results

A high level of responsiveness was demonstrated between baseline and discharge for the EQ-5D-5L index value and VAS with effects sizes and SRM values greater than 0.7. Between discharge and three-month follow-up there was a decline in EQ-5D-5L

scores with negative responsiveness values found for the index value and VAS. Participants predominantly prioritised occupational problems in the self-care domain of the COPM. The COPM demonstrated a high level of responsiveness between baseline and three-month follow-up. Both outcome measures demonstrated a decline in responsiveness by six-month follow-up.

Conclusion

The EQ-5D-5L and COPM are responsive to improvement in HRQOL and occupational performance respectively, following stroke rehabilitation. The reduction in HRQOL post-discharge requires further investigation with important implications for stroke care in the community.

5.2 Chapter overview

This chapter was undertaken to examine the responsiveness of the PROMs which were included in the pilot RCT. This chapter sought to determine the suitability of the outcome measures and the assessed time intervals through examination of this psychometric property. Appendix 1 details the researcher team members involved in the study.

5.3 Introduction

The previous chapter examined the responsiveness of activity-level outcome measures; the FIM and gWMFT. The FIM and gWMFT were found to be responsive outcome measures suitable for inclusion in a pilot RCT examining the effectiveness of mirror therapy in a sub-acute stroke population. The variations in responsiveness found across the time intervals investigated highlighted the utility of the outcome measures in the initial months' post stroke. Further research is needed to consider the external responsiveness of the outcome measures and their ability to capture meaningful change using a range of external criterions including patient-reported and clinician-reported and accepted clinical ratings of change.

Outcome measures conducted solely from the perspective of clinicians, are limited in their ability to capture meaningful and relevant outcomes for stroke survivors (Black 2013). In examining the effectiveness of healthcare interventions, the perspectives of stroke survivors themselves offer meaningful insights; these are captured through the use of PROMs (Devlin and Appleby 2010). Patient-reported outcome measures are validated assessments which often take the form of the individual's perceptions of their functional status or health-related quality life (HRQOL) and have gained prominence

in recent years (Dawson *et al*. 2010). Since April 2009, the National Health Service in England has required the application of PROMs for all unilateral hip and knee replacements, groin hernia surgery and varicose vein surgery (Department of Health 2008), with promotion of their expansion in a governmental white paper report (Department of Health 2010). The International Consortium of Health Outcomes Measurement recommend the inclusion of PROMs as part of a standard set of person-centred outcome measures for individuals with stroke (Salinas *et al*. 2016). Overall, there has been an increase in the use of PROMs in clinical trials (Vodicka *et al*. 2015). However, between 2002 and 2016 less than 25% of registered RCTs in the area of stroke included PROMs (Price-Haywood *et al*. 2019). Patient perspectives are essential to capture the complexity of outcomes associated with stroke and essential to the delivery of person-centred care (Reeve *et al*. 2013).

As part of the pilot RCT which investigated the effectiveness of mirror therapy, two PROMs were included: the EQ-5D-5L, a measure of HRQOL and the COPM, a person-centred evaluation of occupational performance. Health-related quality of life is considered a multidimensional construct, encompassing how the physical, social and psychological domains of an individual's life are affected by disease (Carod-Artal and Egido 2009). The EQ-5D-5L is commonly used across stroke trials (Dyer *et al*. 2010) and is recommended by NICE guidelines in the assessment of health technologies (NICE 2014). Adapted from the EQ-5D-3L, the EQ-5D-5L consists of five domains: mobility, self-care, usual activities, pain and anxiety (EuroQol Research Foundation 2019) (Appendix 11). Responses to each domain comprise five levels of severity: no problems, slight problems, moderate problems, severe problems and extreme problems. In contrast, the EQ-5D-3L consists of three possible responses: no problems, some problems and extreme problems. The EQ-5D-5L was designed to

overcome ceiling effects and increase the measure's ability to discriminate between different health states, particularly when differentiating between milder health states (Janssen *et al*. 2013).

An objective of the pilot RCT was to determine the cost-effectiveness of delivering mirror therapy to a sub-acute stroke population, using the EQ-5D-5L to determine quality-adjusted life years. Responsiveness is an integral property to ensure the EQ-5D-5L is able to accurately measure change and ensure robust findings for the economic evaluation of mirror therapy. Responsiveness of the EQ-5D-3L has been assessed in stroke survivors undergoing inpatient rehabilitation in Taiwan, in the sub-acute phase of stroke (Lu *et al*. 2016). Using multiple analyses of responsiveness, study authors reported moderate responsiveness to change, with a SRM of 0.74, an ES of 0.76, and a significant difference between admission and discharge scores using paired t-tests (Lu *et al*. 2016). Where the responsiveness of the EQ-5D-5L has been assessed, none of the studies investigated the inpatient rehabilitation period for sub-acute stroke survivors (Golicki *et al*. 2015a; Chen *et al.* 2016). Chen *et al*. (2016) assessed responsiveness between baseline and follow-up for both sub-acute and chronic stroke survivors and Golicki *et al*. (2015a) assessed responsiveness following acute admission for stroke and four months later.

The final patient-reported outcome measure used in the pilot RCT was the COPM, a client-centred, occupation-focused outcome measure (Law *et al*. 2014). As an individualised measure, occupational performance issues are defined and evaluated by the participants. The COPM has been used to capture the occupational priorities of stroke survivors at varying time points post-stroke (Almhdawi *et al*. 2016; Schiavi *et al*. 2016; Kessler *et al*. 2017) and with varying interventions including: occupational

performance coaching (Kessler *et al*. 2017), home rehabilitation versus clinic-based rehabilitation (Hsieh *et al*. 2018) and expanded constraint-induced movement therapy (Uswatte *et al*. 2018). Psychometric evaluation of the COPM has been completed across patient groups and settings (Cup *et al*. 2003; Carswell *et al*. 2004; Eyssen *et al*. 2011; Thyer *et al*. 2018). However, examination of the responsiveness of the COPM within a stroke population remains limited: there is wide heterogeneity in client groups evaluated, such as the recruitment of mixed populations from neurological rehabilitation departments, sub-acute rehabilitation settings and amongst community-dwelling older adults (Bodiam 1999; Wressle *et al*. 1999; Chen *et al*. 2002; Eyssen *et al*. 2011; Tuntland *et al*. 2016; Roe *et al*. 2019). Additionally, there is heterogeneity in the methods used to determine responsiveness of the COPM (Bodiam 1999; Wressle *et al*. 1999; Chen *et al*. 2002; Eyssen *et al*. 2011; Tuntland *et al*. 2016; Roe *et al*. 2019). While it is acknowledged that there is no accepted method for assessing responsiveness (Norman *et al*. 2007), multiple analyses are often used within and across studies with the inability to draw comparisons across the different analyses used (Husted *et al*. 2000; Terwee *et al*. 2003).

A recent study highlighted that while clinicians prioritise person-centred and goal attainment outcome measures, encapsulated by the COPM, their use in RCTs remains small (Duncan Millar *et al*. 2019). The COPM is an outcome measure designed to capture change across time and assessment of its responsiveness is essential to ensure sensitive, accurate results. This would support the use of the COPM in clinical trials and clinical practice and help meet the priorities of relevant stakeholders in reflecting important outcomes for patients, carers and clinicians.

In this study the same technique described in Chapter 4 was implemented to assess responsiveness of the EQ-5D-5L and the COPM. Pilot trials are recommended prior to the implementation of a main trial to ensure study procedures are feasible and robust. This study supports the objectives of pilot trials advocated by the Medical Research Council related to appropriate outcome measurement (Lancaster *et al*. 2004). As part of this process, the ability of the PROMs to asses change in the time periods implemented as part of the pilot RCT were investigated.

5.4 Aim and objectives

5.4.1 Aim

The aim of this study was to investigate the responsiveness of the PROMs completed as part of the pilot RCT.

5.4.2 Objectives

The objectives of the study were:

1. To assess the internal and external responsiveness of the EQ-5D-5L between baseline and discharge; baseline and three-month follow-up; baseline and six-month follow-up; discharge and three-month follow-up and three- and six-month follow-up for the full sample, intervention and control groups.
2. To describe the occupational goals prioritised by sub-acute stroke survivors at baseline using the COPM.
3. To assess the internal and external responsiveness of the COPM between baseline and three-months follow-up; baseline and six-month follow-up and between three- and six-months follow-up for the full sample, intervention and control groups.

5.5 Method

5.5.1 Study design

This study was a clinical measurement study, completed as part of a pilot RCT examining the effectiveness of mirror therapy in sub-acute stroke. The psychometric properties of the outcome measures assessed were internal and external responsiveness, as described in Chapter 4.

5.5.2 Participants

Participants were those recruited to the same pilot RCT as reported in Chapter 4 (n=40). Following completion of the gWMFT and FIM at baseline, participant 12 became unwell and did not complete the EQ-5D-5L and the COPM. Subsequently participant 12 was withdrawn from the pilot RCT, and their data was omitted from the analysis of this study (n=39).

5.5.3 Ethical approval

The same ethical approval governing the pilot RCT, as reported in Chapter 3, also applied to this study.

5.5.4 Outcome measures

5.5.4.1 EQ-5D-5L

The EQ-5D-5L is a generic measure of HRQOL (EuroQol Research Foundation 2019) designed to be self-completed by each participant. The EQ-5D-5L consists of a questionnaire based on the areas of mobility, self-care, usual activities, pain/discomfort and anxiety/depression, which are scored according to five levels of severity on a Likert scale (Appendix 11**)**.

A status value is generated based on scores across each dimension, for example a score of 1, 1, 1, 1, 1 represents optimal health and no problems reported, and a score of 5, 5, 5, 5, 5 represents severe problems in all areas. These status values can be converted into a single country-specific index value.

Index values are formed using the preferences from country-specific population responses, generating a number between zero (death) and one (optimum health), scores less than zero are possible and represent a state worse than death. The index value is commonly used in economic evaluations and allows for comparison of values across trials and conditions (Dyer *et al*. 2010). EQ-5D-5L value sets have been developed in several countries including Germany, Spain, Indonesia, Japan and China, with work ongoing for many countries globally (EuroQol 2020). While a value set has been developed for England (Devlin *et al*. 2016), to date there have been none developed for the UK.

Currently there is no EQ-5D-5L index set available for the UK and an interim value set was developed by Van Hout *et al*. (2012). This crosswalk method was devised whereby health states from the EQ-5D-5L were mapped onto those for the EQ-5D-3L, allowing index values to be derived from health states previously developed for the EQ-5D-3L (Van Hout *et al*. 2012). This study involved 3691 participants across Denmark, the Netherlands, Poland and the UK who completed the EQ-5D-5L and the EQ-5D-3L. From participant responses a relationship between the two versions of the outcome measure was developed which enabled country-specific value sets to be generated for the EQ-5D-5L. This is the technique currently recommended by NICE when generating index value scores for the EQ-5D-5L in the UK (NICE 2018).

The EQ-5D-5L also includes a visual analogue scale (VAS) which was used to measure each participant's perceived level of health (Appendix 11), from worst health imaginable (0) to best health imaginable (100).

5.5.4.2 Canadian Occupational Performance Measure

The COPM is a client-centred, occupation-focused, goal setting tool, commonly used within occupational therapy practice (Law *et al*. 2014). Developed from the Canadian Model of Occupational Performance (Townsend *et al.* 2002), now the Canadian Model of Occupational Performance and Engagement (Townsend and Polatajko 2007), occupational performance is presented as the dynamic interaction between the person, environment and meaningful occupation (Law *et al*. 2014). A semi-structured interview format allows individuals to prioritise occupational performance issues in the domains of self-care, productivity and leisure. Routinely, but not exclusively, used by occupational therapists (Law *et al*. 2014), the COPM identifies up to five occupations the individual wishes to enhance their performance and satisfaction in, across the three domains.

Participants are asked to identify activities they are expected to do, want to do or need to do in the areas of self-care, productivity and leisure. Where possible the activities reported were classified, whereby washing, dressing and showering were classified under the single category of washing/dressing. Activities as reported by the participants are listed in Appendix 12.

5.5.5 Scoring and procedure

Scoring was completed by the same research occupational therapist for both outcome measures. Training on administration and scoring of the outcome measures was

completed prior to study commencement. Participants were assessed at baseline and prior to discharge in a private research room on hospital premises. Assessments completed at three- and six-month follow-up generally took place in participants own homes.

The paper version of the EQ-5D-5L was completed at baseline, discharge, three-month and six-month follow-up. The EQ-5D-5L was filled in by the assessing therapist where a participant's dominant hand affected by the stroke impeded their ability to write. The VAS aspect of the EQ-5D-5L was also completed. The VAS consists of a 20cm vertical ruler numbered from zero to 100. Either end of the vertical scale is labelled 'The worst health you can imagine' and 'The best health you can imagine' at the numbers zero and 100 respectively. Participants were required to mark an 'x' on the vertical scale at the number corresponding to their general health as they perceived it and then write the number in a box beside the scale (Appendix 11).

The COPM was assessed at baseline and at three- and six-month follow-up, see flowchart for details of the assessment procedure (Appendix 10). Previously identified activity goals were reviewed at three- and six-months. Occupational performance issues reported by participants were rated in terms of their importance, and the top five were identified as key occupational performance problems. Participants then rated their level of perceived performance for each task from one (not able to perform) to ten (perform extremely well), and how satisfied they were with this performance from one (not satisfied) to ten (extremely satisfied), using a Likert scale as a guide. Once the participant assigned scores, these were added together to generate a total score for performance and a total score for satisfaction, and then divided by the number of

activities reported to generate a mean performance score and a mean satisfaction score.

5.5.6 Data analysis

Participant results were collated into an Excel spreadsheet. Following this the data was cleaned and entered into SPSS. Data analysis was completed using SPSS Version 25.0 (SPSS Inc. Chicago) and Microsoft Excel 2016. Although the aim of the pilot study was not powered to detect statistical significance, analyses were completed for the total sample, and for the intervention and control groups separately. Descriptive statistics were used to report characteristics of the sample: gender, age, side of hemiplegia and days post-stroke. Frequency distributions for each dimension of the EQ-5D-5L were reported and the distribution of problems identified in the COPM were also reported.

Means and standard deviations were reported for the EQ-5D-5L index values, EQ-5D VAS, COPM performance and COPM satisfaction scores at baseline. A linear mixed model was completed for the EQ-5D-5L and COPM, as described in Chapter 4, due to decreases in participant numbers at each assessment point. Estimated marginal means and standard errors were reported for discharge (where applicable), three-month and six-month follow-up, controlling for baseline values.

5.5.6.1 Floor and ceiling effects

The status values for the EQ-5D-5L were checked for floor and ceiling effects. These effects were considered present when 15% or more of the sample achieved a score of 1, 1, 1, 1, 1 indicating optimum health or a score of 5, 5, 5, 5, 5 indicating extreme difficulties (McHorney and Tarlov 1995).

5.5.6.2 Responsiveness

Internal Responsiveness

As reported in Chapter 4, responsiveness was assessed using the standardised ES and SRM. The ES was computed by dividing the mean difference in scores by the standard deviation of the first assessment for each time period (Husted *et al*. 2000). The SRM was computed by dividing the mean difference in scores by the standard deviation of the mean change (Liang *et al*. 1990; Husted 2000). This was completed for the total sample and for the intervention and control groups for each time interval. Positive values were viewed as improvements in HRQOL and occupational performance for the EQ-5D-5L and COPM respectively. This was completed for changes in scores on the EQ-5D-5L between baseline and discharge; discharge and three-month follow-up; three- and six-month follow-up. Responsiveness was also evaluated between baseline and three-month and baseline and six-month follow-up. Standardised response mean values and ES were also completed for COPM performance and satisfaction scores for all time points, from baseline to three-month follow-up; three- and six-month follow-up and baseline and six-month follow-up. These were completed for those participants who were assessed at both time points for each time interval, for the total sample, intervention group and control group.

External Responsiveness

The external criterion reported in Chapter 4 was also used in the current study. The following question was answered as part of the exit questionnaire, which was completed prior to discharge: “In your opinion do you feel that your occupational therapy sessions have been of any benefit with regard to the return of movement to your affected arm?”, and optional answers were ‘very beneficial’, ‘quite beneficial’, ‘of little benefit’ or ‘of no benefit’. Responses were numbered, one to four, with one

representing ‘no benefit’ to four representing ‘very beneficial’. This was transformed into a dichotomous scale, to denote patient-reported rating of change, for assessing external responsiveness. The external criterion was administered prior to discharge and as such it was not seen as appropriate to apply this to the COPM which was not administered at discharge.

Associations between the external criterion and EQ-5D-5L index values and EQ-5D VAS scores were competed initially to determine the strength of relationship. Correlations above 0.3 were considered an adequate association (Revicki *et al*. 2008). Using Spearman’s Rho, the correlations between the external criterion and the differences in scores from baseline to discharge for the EQ-5D-5L index values and EQ-5D VAS were less than 0.2 and thus were not adequate to analyse external responsiveness.

5.6 Results

Thirty-nine participants completed the EQ-5D-5L and COPM at baseline. Participant characteristics and scores for the EQ-5D-5L index values, EQ-5D VAS and COPM at baseline are shown in Table 5.1. Participants were approximately 15 days post-stroke and more males than females were recruited. The number of males in either treatment group almost doubled that of the number of females recruited.

Table 5.1 Participant characteristics and EQ-5D-5L index values and visual analogue scale scores at baseline

	Total Sample n=39	**Intervention n=19**	**Control n=20**
Age in years, mean (SD)	71.13 (9.32)	72.58 (9.78)	69.75 (8.9)
Days post-stroke, mean (SD)	15.31 (8.73)	15.95 (9.29)	14.70 (8.36)

Gender			
Male, n (%)	26 (66.7%)	13 (68.4%)	13 (65%)
Female, n (%)	13 (33.3%)	6 (31.6%)	7 (35%)
Side of hemiplegia			
Left, n (%)	24 (61.5%)	10 (52.6%)	14 (70%)
Right, n (%)	15 (38.5%)	9 (47.4%)	6 (30%)
EQ-5D-5L index value			
Mean (SD)	0.16 (0.31)	0.16 (0.35)	0.15 (0.28)
EQ-5D-5L VAS			
Mean (SD)	39.72 (27.92)	39.98 (31.93)	39.75 (24.36)
COPM Performance			
Mean (SD)	2.38 (1.31)	2.39 (1.51)	2.37 (1.14)
COPM Satisfaction			
Mean (SD)	2.47 (1.64)	2.57 (1.71)	2.38 (1.61)

Abbreviations: SD, standard deviation; %, percentage; COPM, Canadian Occupational Performance Measure

Participant two and four did not identify any occupational goals at six-month follow-up for the COPM, and as such no data was available at this time point. The sample size at each assessment point for the EQ-5D-5L and COPM are detailed in Table 5.2 and Table 5.3 respectively.

5.6.1 EQ-5D-5L descriptive statistics

The estimated marginal means according to time and treatment group, controlling for baseline scores are reported in Table 5.2 for the EQ-5D-5L index values and EQ-5D VAS at discharge, three-month and six-month follow-up.

Table 5.2 Estimated marginal means (EMM) and standard error (SE) values for the EQ-5D-5L for the total sample, intervention and control groups, controlling for baseline values

	Discharge			**Three-month**			**Six-month**		
	Full sample n=34	**Intervention n=17**	**Control n=17**	**Full sample n=31**	**Intervention n=17**	**Control n=14**	**Full sample n=25**	**Intervention n=15**	**Control n=10**
EQ-5D-5L Index value									
EMM (SE)	0.49 (0.04)	0.58 (0.06)	0.41 (0.06)	0.43 (0.04)	0.51 (0.06)	0.36 (0.06)	0.46 (0.05)	0.56 (0.06)	0.35 (0.07)
EQ-5D-5L VAS									
EMM (SE)	63.14 (3.86)	64.45 (5.16)	61.82 (5.29)	59.69 (4.04)	67.78 (5.16)	51.58 (5.81)	68.02 (4.56)	71.01 (5.59)	64.94 (6.74)

Abbreviations: EMM, estimated marginal mean; SE, standard error; VAS, visual analogue scale.

The proportion of responses to the five dimensions of the EQ-5D-5L are reported in Figure 5.1 to Figure 5.5 for the whole sample, Figure 5.6 to Figure 5.10 for the intervention group and Figure 5.11 to Figure 5.15 for the control group. At baseline just over 50% of the total sample reported they were unable to mobilise and engage in usual activities (Figure 5.1 and Figure 5.3). By discharge, 5% of the total sample reported they were unable to mobilise, with 30% reporting moderate difficulties with mobility (Figure 5.1). By discharge 14% of the total sample reported they were unable to complete usual activities, with 14% reporting severe problems in this area (Figure 5.3).

Figure 5.1 Proportion of responses by level of severity for the mobility dimension of the EQ-5D-5L for the total sample at baseline, discharge, three months and six months

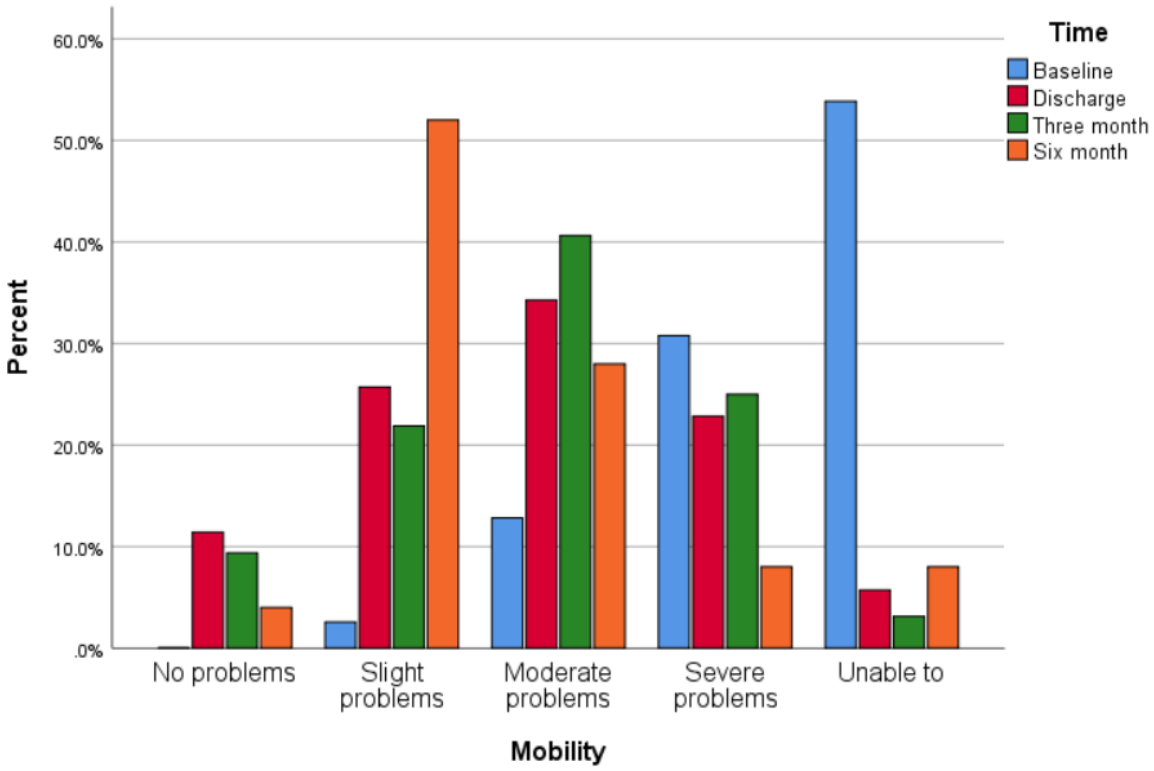


Figure 5.2 Proportion of responses by level of severity for the self-care dimension of the EQ-5D-5L for the total sample at baseline, discharge, three months and six months

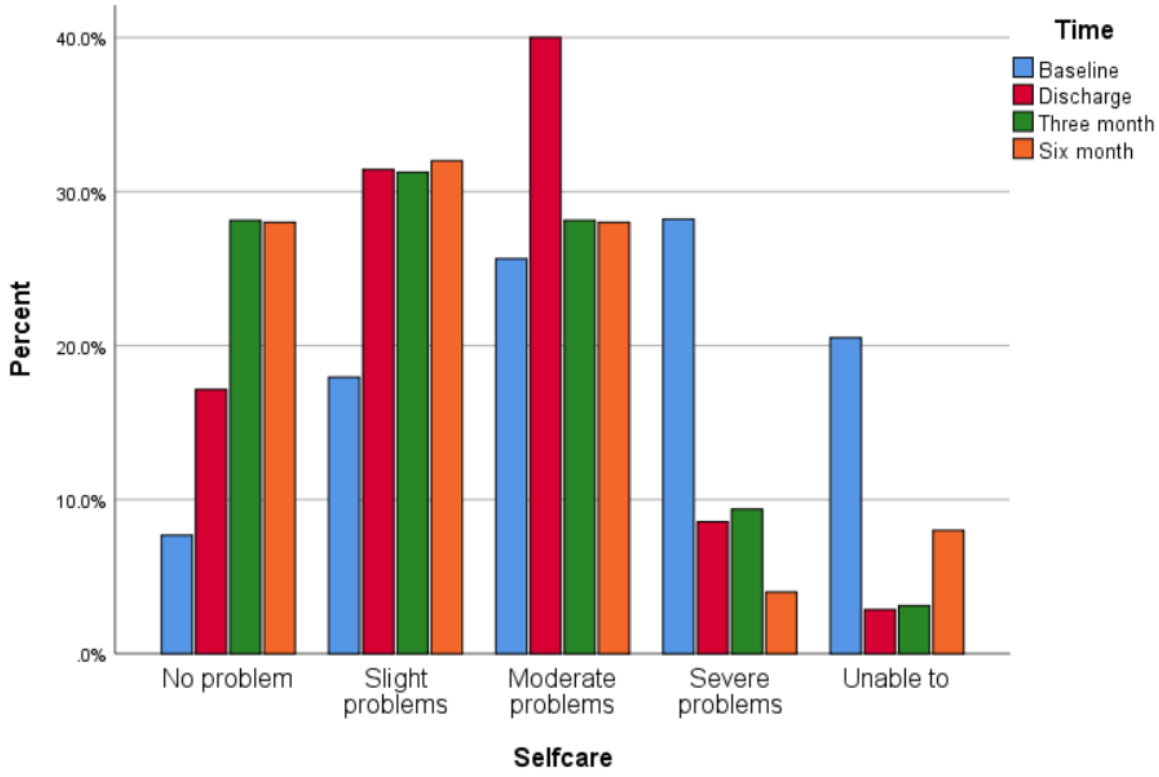


Figure 5.3 Proportion of responses by level of severity for the activity dimension of the EQ-5D-5L for the total sample at baseline, discharge, three months and six months

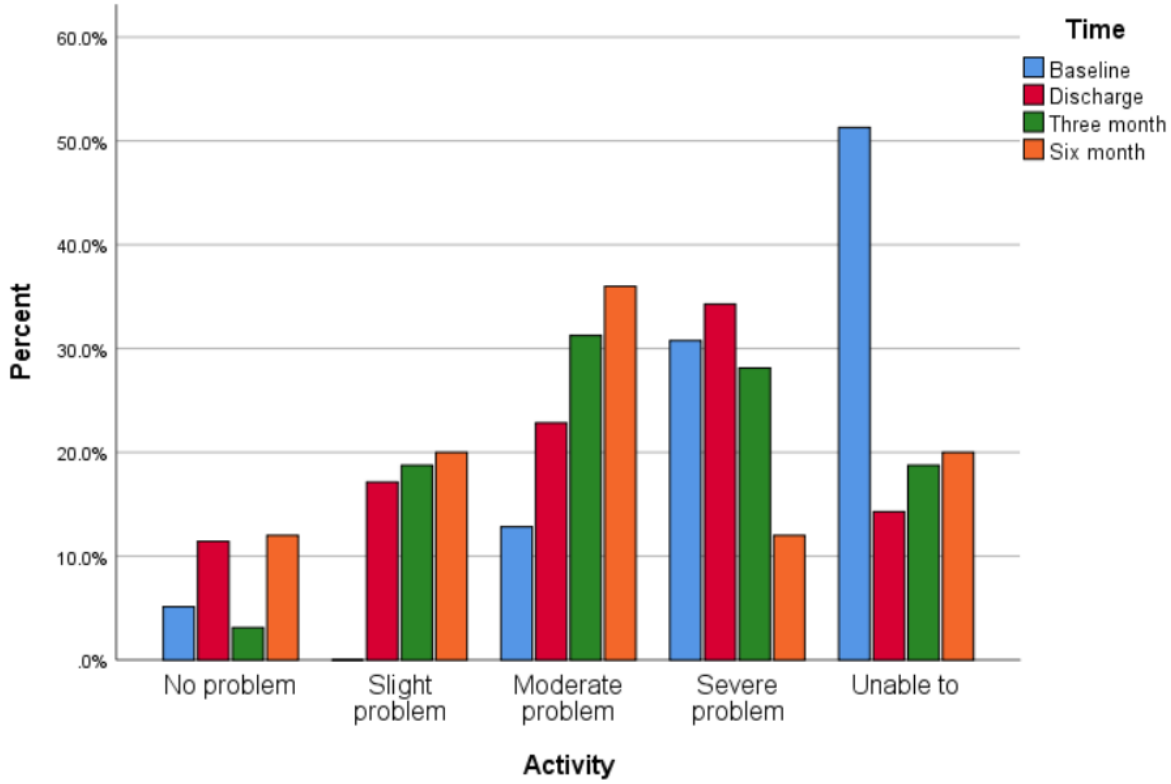


Figure 5.4 Proportion of responses by level of severity for the pain dimension of the EQ-5D-5L for the total sample at baseline, discharge, three months and six months

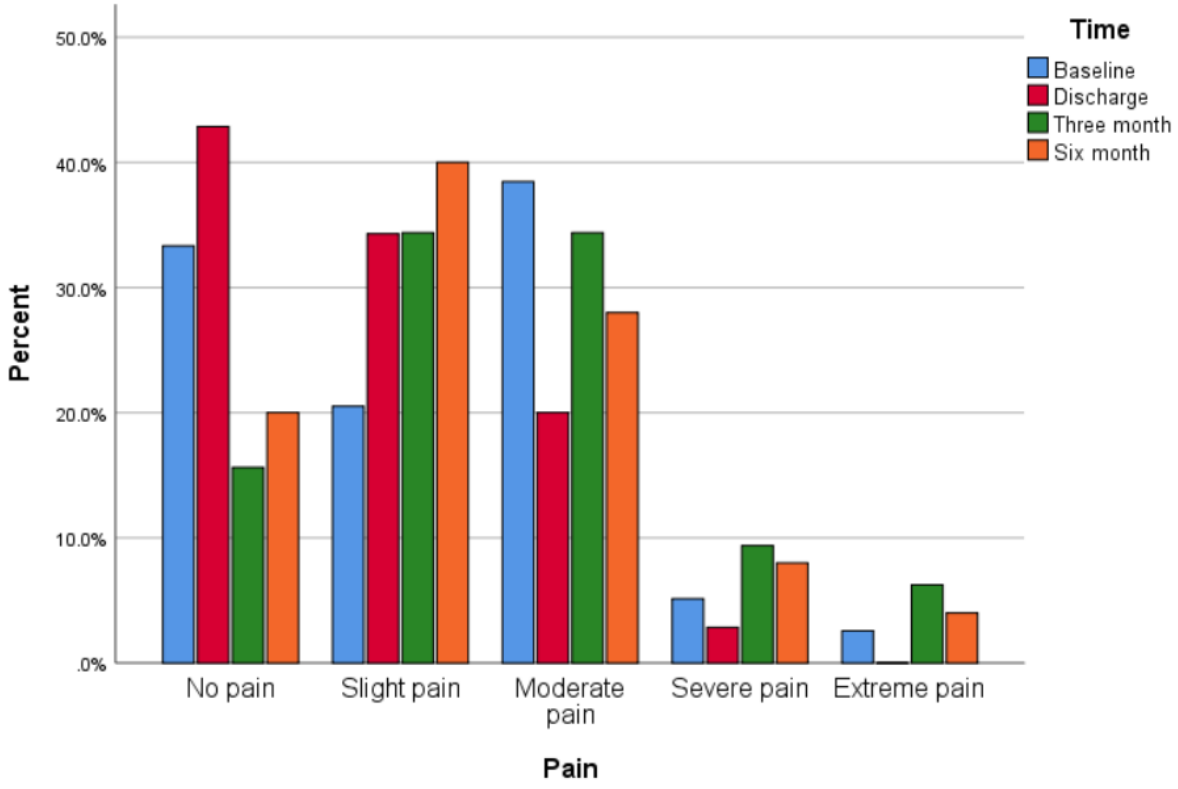
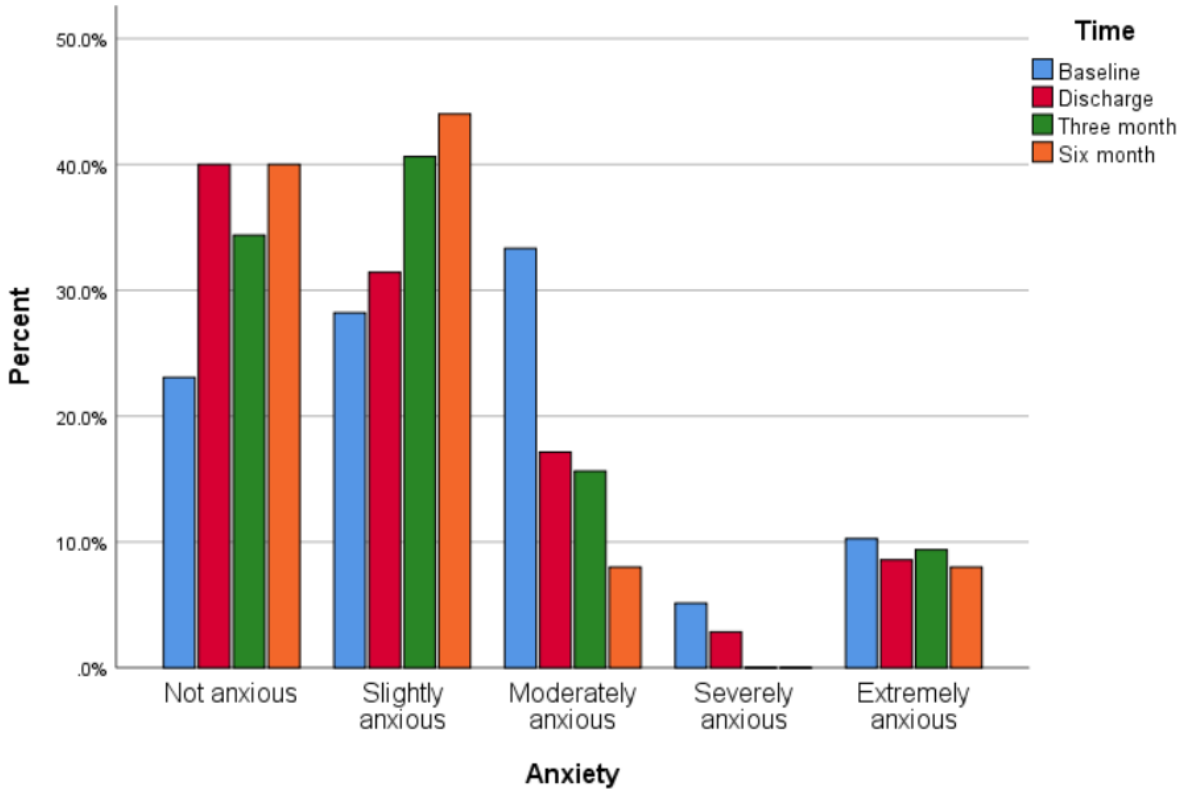


Figure 5.5 Proportion of responses by level of severity for the anxiety dimension of the EQ-5D-5L for the total sample at baseline, discharge, three months and six months



The proportion of responses for the treatment groups demonstrated that most participants reported most difficulties with mobility and usual activities; at baseline 50% of participants reported they were unable to mobilise (Figure 5.6 and Figure 5.11) and just less than 40% reported they were unable to engage in usual activities (Figure 5.8 and Figure 5.14). More participants in the control group reported problems with pain, in comparison to the intervention group; at baseline 42% of participants in the intervention group reported no problems with pain (Figure 5.9) in comparison to 25% of the control group (Figure 5.14). Detailed frequency distributions for the total and treatment groups are in Appendix 13 and Appendix 14 respectively.

Figure 5.6 Proportion of responses by level of severity for the mobility dimension of the EQ-5D-5L for the intervention group at baseline, discharge, three months and six months

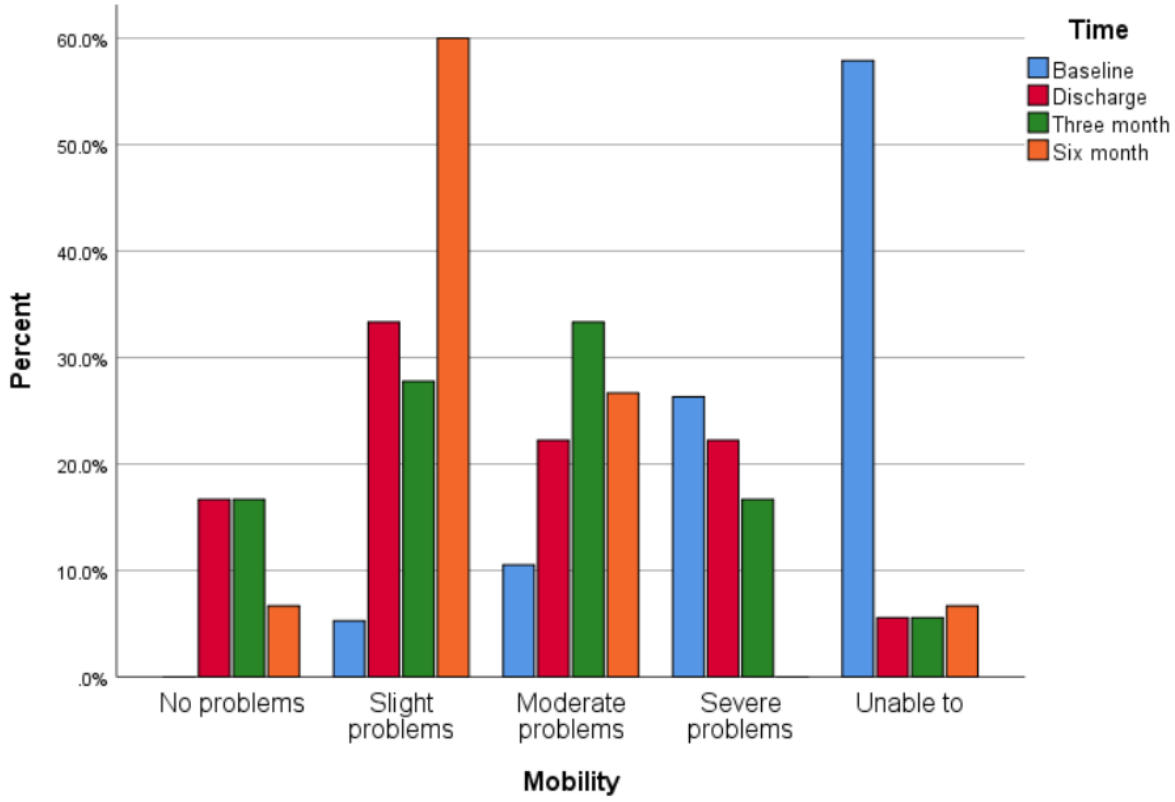


Figure 5.7 Proportion of responses by level of severity for the self-care dimension of the EQ-5D-5L for the intervention group at baseline, discharge, three months and six months

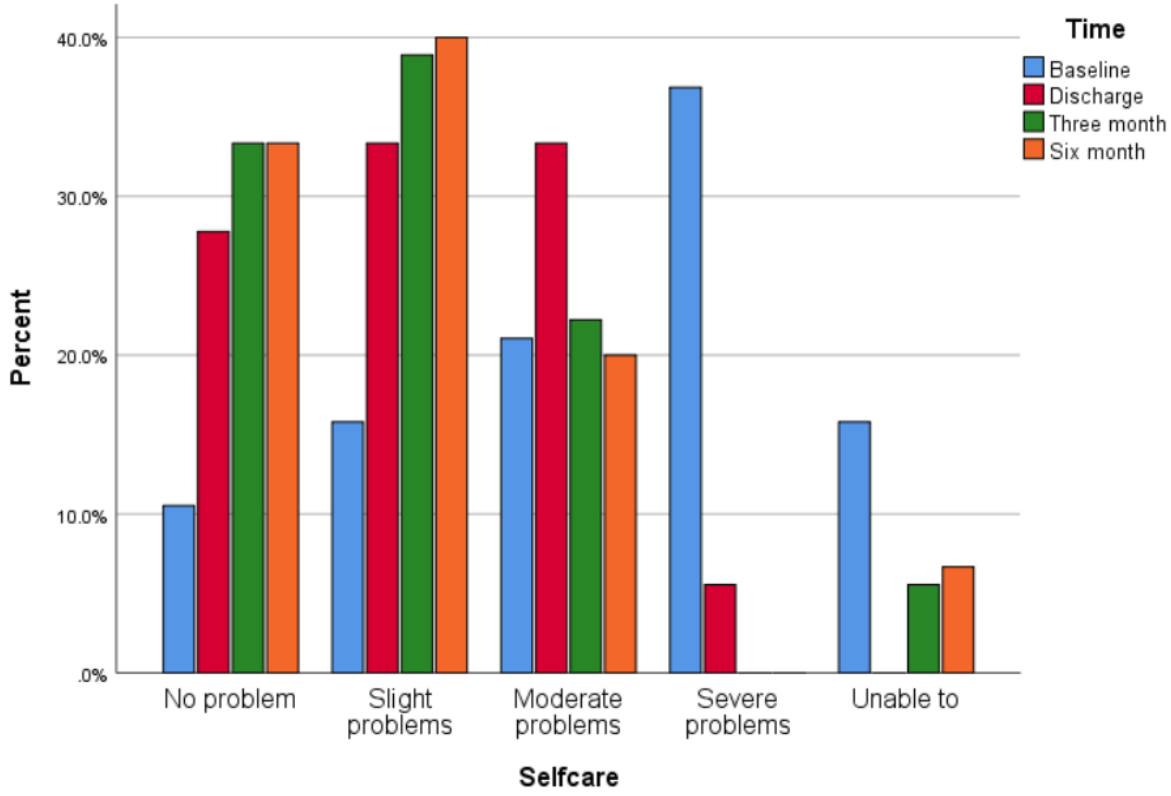


Figure 5.8 Proportion of responses by level of severity for the activity dimension of the EQ-5D-5L for the intervention group at baseline, discharge, three months and six months

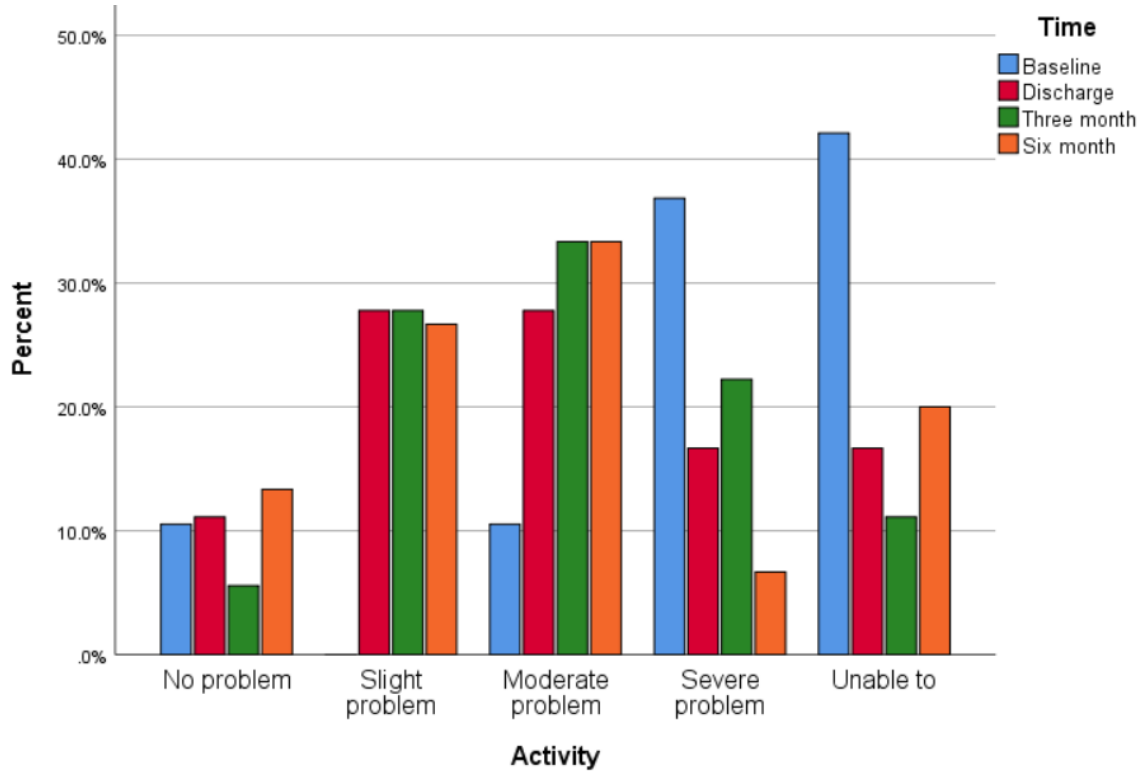


Figure 5.9 Proportion of responses by level of severity for the pain dimension of the EQ-5D-5L for the intervention group at baseline, discharge, three months and six months

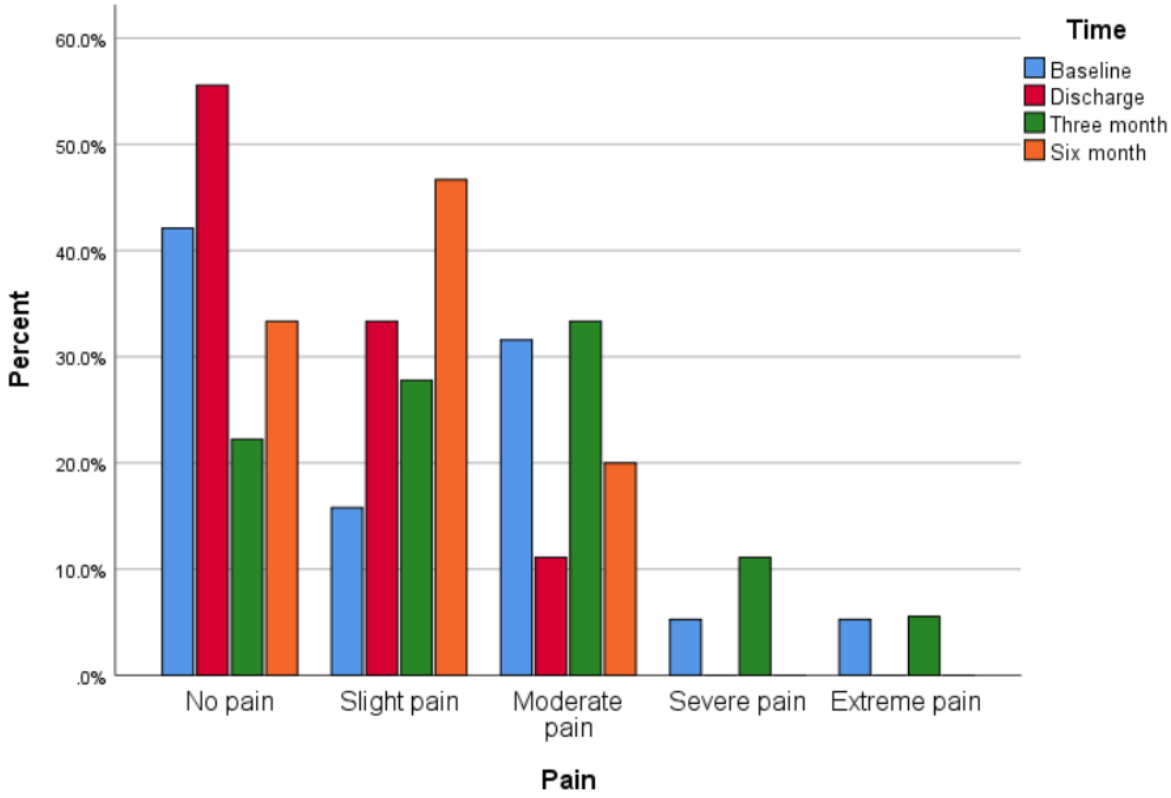


Figure 5.10 Proportion of responses by level of severity for the anxiety dimension of the EQ-5D-5L for the intervention group at baseline, discharge, three months and six months

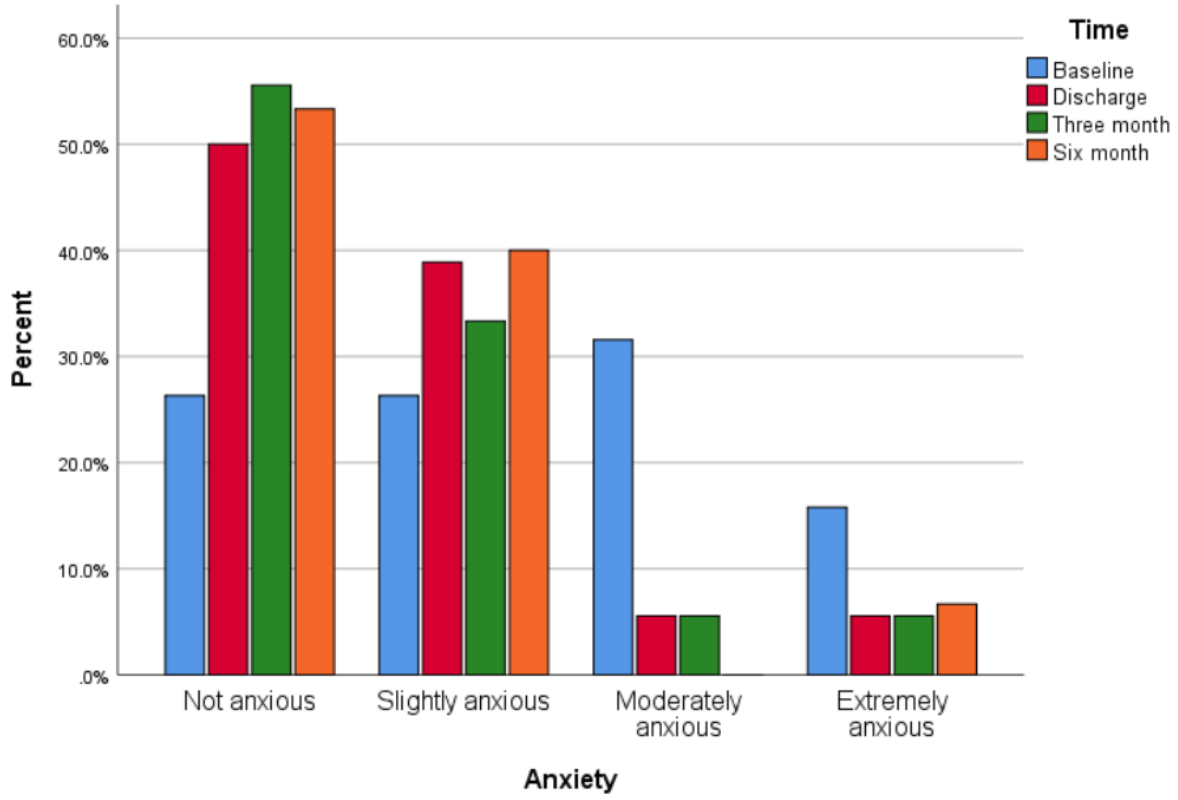


Figure 5.11 Proportion of responses by level of severity for the mobility dimension of the EQ-5D-5L for the control group at baseline, discharge, three months and six months

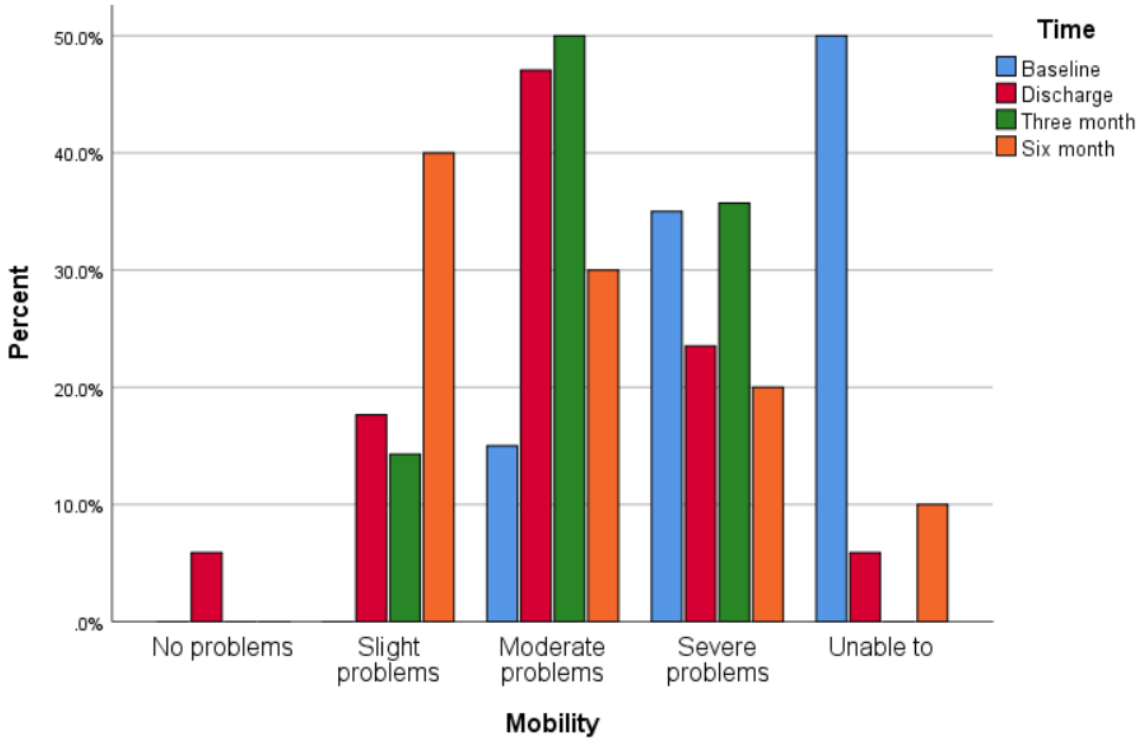


Figure 5.12 Proportion of responses by level of severity for the self-care dimension of the EQ-5D-5L for the control group at baseline, discharge, three months and six months

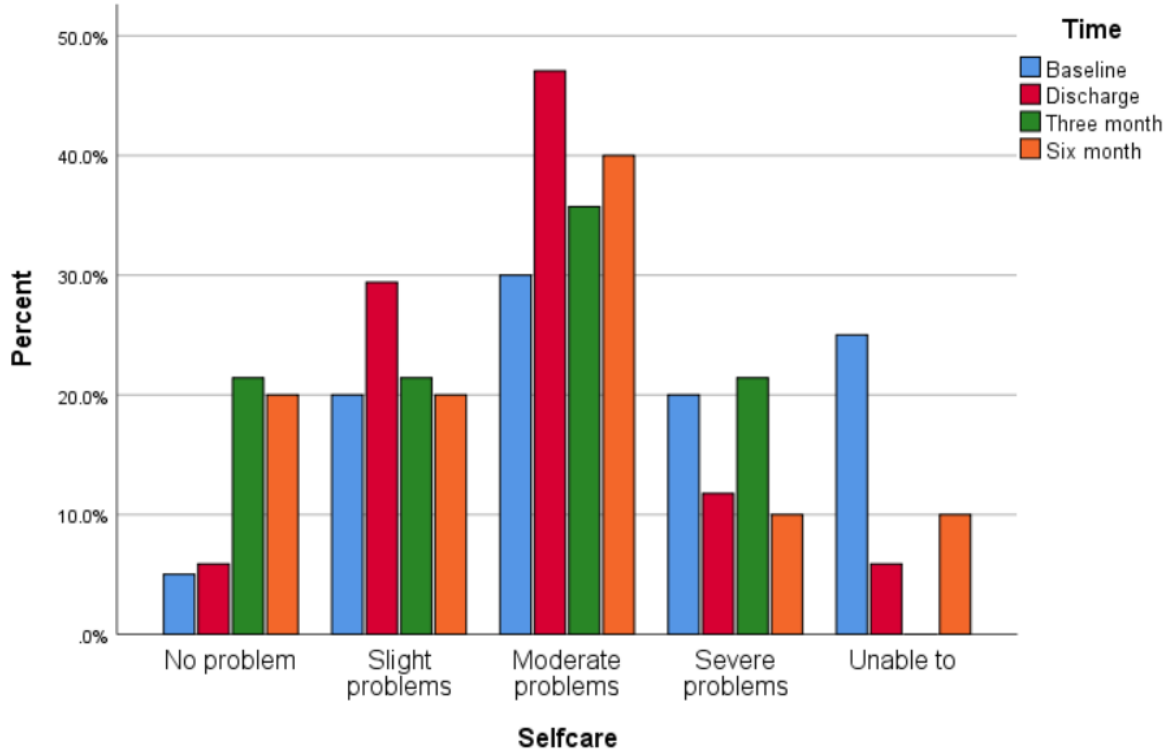


Figure 5.13 Proportion of responses by level of severity for the activity dimension of the EQ-5D-5L for the control group at baseline, discharge, three months and six months

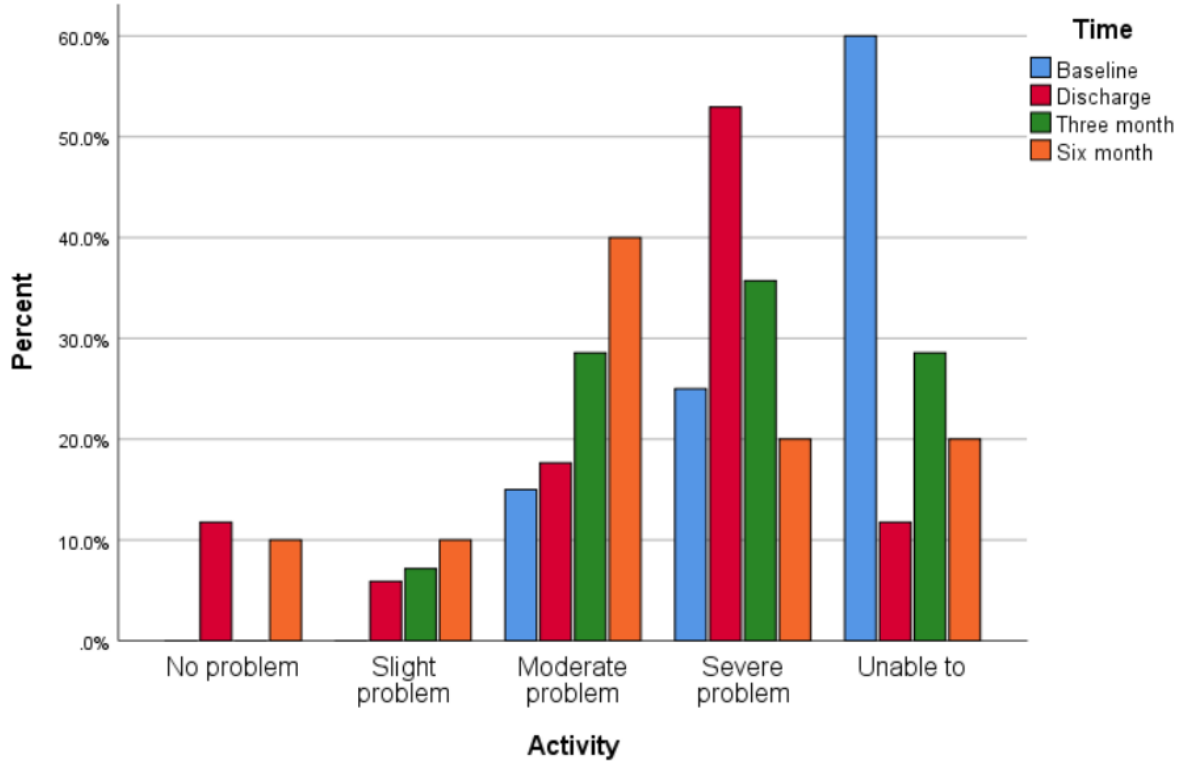


Figure 5.14 Proportion of responses by level of severity for the pain dimension of the EQ-5D-5L for the control group at baseline, discharge, three months and six months

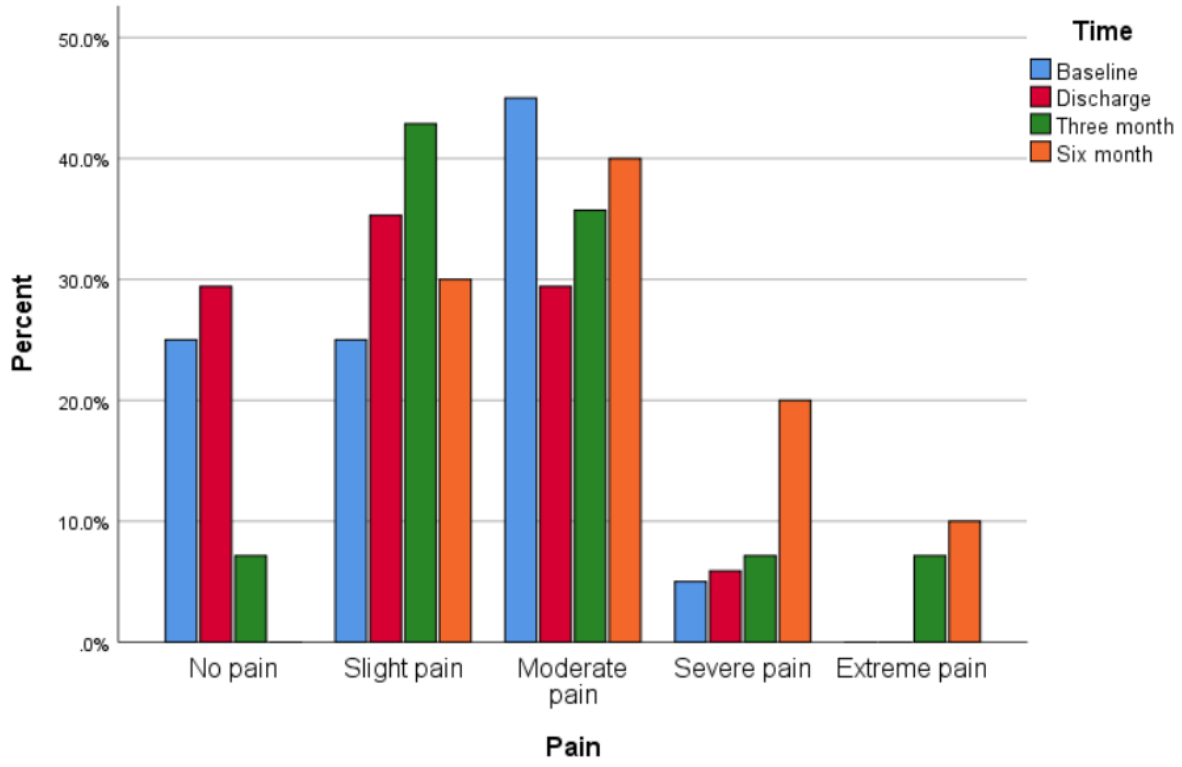
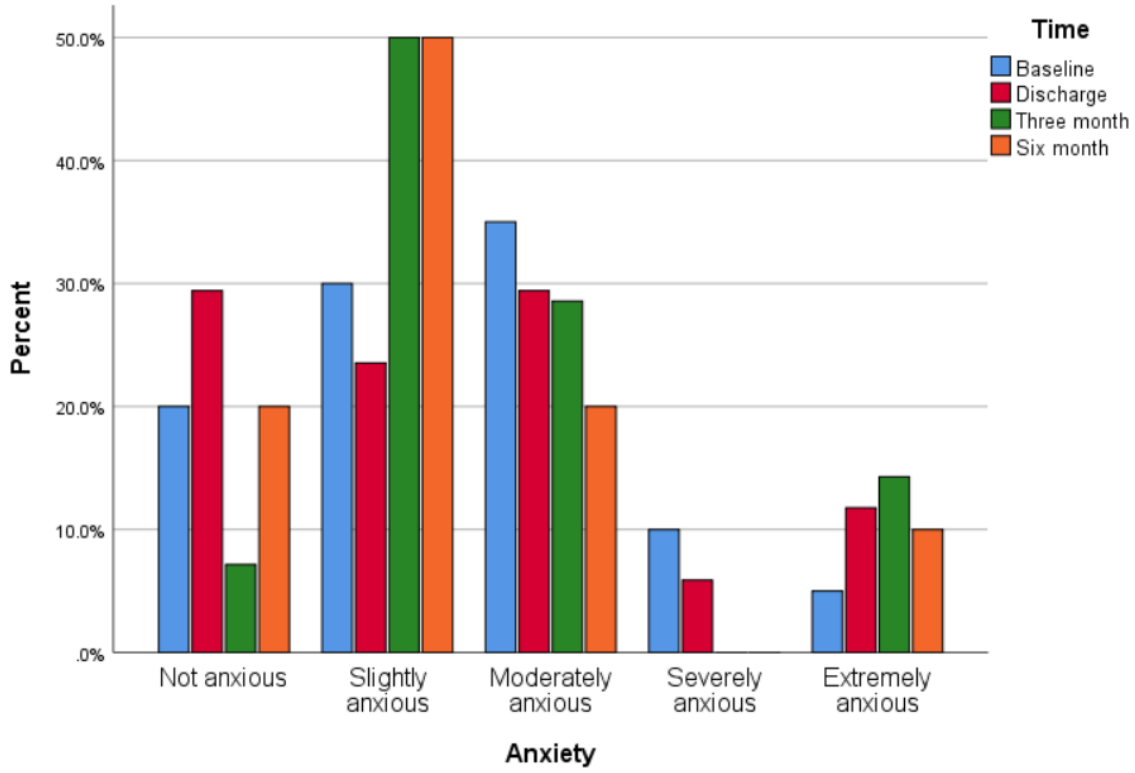


Figure 5.15 Proportion of responses by level of severity for the anxiety dimension of the EQ-5D-5L for the control group at baseline, discharge, three months and six months



5.6.2 Canadian Occupational Performance Measure descriptive statistics

The COPM was completed at baseline and repeated at three- and six-months' follow-up. The mean values for the COPM performance and satisfaction at baseline are reported in Table 5.1, and the estimated marginal means and standard error values are reported in Table 5.3.

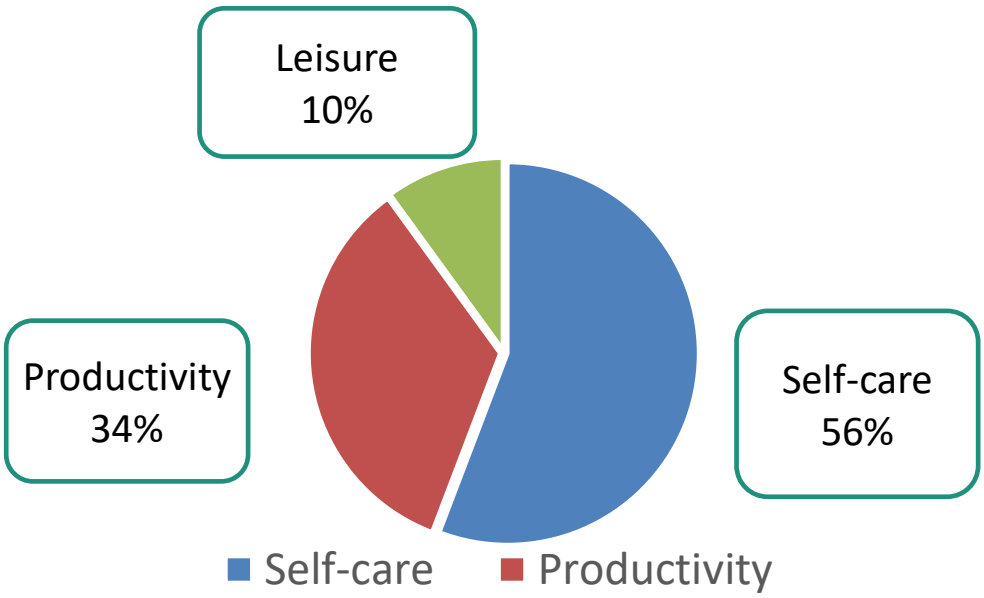
Table 5.3 Estimated marginal means (EMM) and standard error (SE) values for the Canadian Occupational Performance Measure (COPM) for the total sample, intervention and control groups, controlling for baseline values

	Three-month			**Six-month**		
	Full sample n=32	**Intervention n=18**	**Control n=14**	**Full sample n=23**	**Intervention n=14**	**Control n=9**
COPM Performance						
EMM (SE)	4.91 (0.37)	5.08 (0.49)	4.74 (0.55)	5.36 (0.41)	5.9 (0.53)	4.83 (0.64)
COPM Satisfaction						
EMM (SE)	5.66 (0.46)	5.95 (0.60)	5.37 (0.68)	5.87 (0.51)	6.35 (0.65)	5.38 (0.79)

Abbreviations: COPM, Canadian Occupational Performance Measure; EMM, estimated marginal mean; SE, standard error.

The 39 recruited participants prioritised 37 occupations in total (Table 5.5). Participants predominantly identified problems in the self-care domain (n=106), followed by productivity (n=46), then leisure (n=42) (Figure 5.16).

Figure 5.16 Occupational performance problems identified by participants using the Canadian Occupational Performance Measure



The most frequently identified problems were washing and dressing, followed by walking and driving which all fell under the self-care domain (Table 5.4). Housework and cooking were most frequently identified under the productivity domain, followed by caring responsibilities and work. Activities related to socialising with friends and family were more frequently identified by participants under the leisure domain, followed by taking part in hobbies. Hobbies encapsulated a broad range of activities meaningful for participants such as playing the guitar, knitting, woodwork, flower arrangement and dancing.

Table 5.4 Participant-identified occupational problems at baseline

Category		**Frequency**
Self-care	Washing/dressing	23
	Walking	21
	Driving	17
	Hygiene	15
	In/out of bed	11
	Grooming	5
	Stairs	5
	Eating	2
	Public transport	2
	Opening containers	2
	Mobility scooter	1
	Shopping	1
	Making cup of tea	1
Productivity	Cleaning	9
	Cooking	7
	Grocery shopping	5
	Caring responsibilities	5
	Paid/unpaid work	8
	Looking after animals	4
	Laundry	3
	Gardening	2
	Role as mother/grandmother	1
	Choosing meals	1
	Use of eBay	1
Leisure	Social outings	8
	Hobbies	8
	Walking dogs	4
	Reading	4
	Gardening	4
	Walking (outside)	3
	Attending church	2
	Watching television	2
	Golf	2
	Social club	2
	Using kindle device	1
	Listening to radio	1
	Writing	1

5.6.3 Floor and ceiling effects

There were no floor or ceiling effects found for the EQ-5D-5L status values. One participant reported the best possible health status (1, 1, 1, 1, 1) at discharge and one

participant reported the worst possible health status (5, 5, 5, 5, 5) at baseline. There were no floor or ceiling effects found for the COPM performance and satisfaction scores.

5.6.4 EQ-5D-5L

Table 5.5 shows the mean differences in scores between all time-points for the full sample and treatment groups for the EQ-5D-5L index values and EQ-5D VAS. Mean differences indicated improvements for the EQ-5D-5L index values and EQ-5D VAS for all time points, except for scores between discharge and three-month follow-up. The greatest difference in scores were demonstrated between baseline and discharge, with similar values found for changes in scores between baseline and three-month, exempting control group scores for the EQ-5D VAS.

5.6.4.1 Internal responsiveness

Results for responsiveness of the EQ-5D-5L and EQ-5D VAS are shown in Table 5.5. The greatest degree of responsiveness was seen between baseline and discharge across all groups, indicated by ES values greater than 0.8, with similar values found using SRM (Table 5.5). The negative ES and SRM values between discharge and three-month follow-up indicated a worsening in HRQOL. Most notable was the SRM value for the control group at this time period at -0.56. Responsiveness between three- and six-month follow-up were mostly minimal, with a greater degree of responsiveness found for the EQ-5D VAS for the control group at 0.56.

The EQ-5D VAS exhibited a mean change of 6.64 (SD=30.9) for the control group between baseline and three-month, in comparison to a mean change of 19.03 (SD=32.84) and 23.74 (SD=37.64) for the full sample and intervention group, respectively. Observation of item level data identified a participant who reported a

score of 85 on the EQ-5D VAS at baseline and a score of 8 at three-months. Removal of this participant led to a mean difference score of 13.08 (SD=20.16) and greater levels of responsiveness with an ES of 0.57 and a SRM value of 0.65 for the control group.

Table 5.5 Mean differences and standardised response mean (SRM) values for the EQ-5D-5L at baseline, discharge, three-month and six-month follow-up

Time point		**Mean difference (SD)**		**ES**		**SRM**	
		EQ-5D-5L	**EQ VAS**	**EQ-5D-5L**	**EQ VAS**	**EQ-5D-5L**	**EQ VAS**
Baseline to discharge	**Full Sample (n=35)**	0.36 (0.34)	23.69 (28.79)	1.14	0.83	1.05	0.82
	Intervention (n=18)	0.44 (0.29)	25.33 (28.27)	1.28	0.79	1.5	0.89
	Control (n=17)	0.27 (0.22)	21.94 (30.1)	0.92	0.87	0.72	0.73
Discharge to three-month	**Full sample (n=31)**	-0.07 (0.22)	-1.84 (25.48)	-0.25	-0.07	-0.31	-0.07
	Intervention (n=18)	-0.07 (0.23)	3.33 (29.95)	-0.27	0.12	-0.31	0.11
	Control (n=13)	-0.07 (0.22)	-9 (16.04)	-0.24	-0.38	-0.3	-0.56
Three-month to six-month	**Full sample (n=25)**	0.03 (0.20)	8.08 (34.09)	-0.28	0.22	0.15	0.24
	Intervention (n=15)	0.05 (0.20)	3 (38.12)	0.18	0.12	0.27	0.08
	Control (n=10)	-0.01 (0.20)	15.7 (27.05)	-0.02	0.53	-0.03	0.58
Baseline to three-month	**Full sample (n=32)**	0.32 (0.36)	19.03 (32.84)	1.09	0.66	0.9	0.58
	Intervention (n=18)	0.37 (0.38)	23.74 (37.64)	1.07	0.89	0.96	0.63
	Control (n=14)	0.26 (0.33)	6.64 (30.9)	0.97	0.27	0.79	0.21

Baseline to six-month	**Full sample (n=25)**	0.37 (0.31)	26.84 (24.36)	1.13	0.86	1.20	1.10
	Intervention (n=15)	0.42 (0.28)	31.07 (28.61)	1.15	0.91	1.47	1.09
	Control (n=10)	0.3 (0.34)	20.5 (15.31)	1.08	0.75	0.87	1.34

Abbreviations: SD, standard deviation; ES, effect size; SRM, standardised response mean; VAS, visual analogue scale.

5.6.5 Canadian Occupational Performance Measure

5.6.5.1 Mean differences

Mean difference values for COPM performance and satisfaction scores for each assessment period are reported in Table 5.6. Mean difference scores at each time period indicated improvement in occupational performance and satisfaction. Scores for the satisfaction sub-scale of the COPM demonstrated the greatest improvement in score between for each time period. Change scores greater than three were found for the total sample, intervention and control groups between baseline and three months. Minimal change was indicated for the total sample, and intervention and control groups between three- and six-month follow-up, as indicated by change scores less than one.

5.6.5.2 Internal responsiveness

The COPM was most responsive between baseline and three-month follow-up; the ES was large for COPM performance scores (ES>1) and COPM satisfaction scores (ES>2) for the total sample and for the intervention and control groups (Table 5.6). Standardised response mean values between baseline and three-month follow-up were similar to those reported for the ES for COPM performance scores. When compared to ES between baseline and three-month follow-up, lower SRM values were found for COPM satisfaction scores for the total sample (SRM=1.13), intervention group (SRM=1.07) and control group (SRM=1.21) indicating greater variance in change scores (Table 5.6). The COPM was least responsive between three-month and six-month follow-up, with ES and SRM values less than 0.4 indicating minimal change. The COPM performance scores were more responsive with an ES of 0.32 found for the total sample (SRM=0.37) in comparison to COPM satisfaction scores,

with an ES of 0.15 found for the total sample (SRM=0.2), between three-month and six-month follow-up.

Table 5.6 Mean differences and standardised response mean (SRM) values for the Canadian Occupational Performance Measure (COPM) performance and satisfaction scores

		Mean difference (SD)		ES		SRM	
		COPM Perf	**COPM Satis**	**COPM Perf**	**COPM Satis**	**COPM Perf**	**COPM Satis**
Baseline to three-month	**Full sample (n=32)**	2.72 (1.95)	3.36 (2.97)	1.39	2.06	1.39	1.13
	Intervention (n=18)	2.85 (1.71)	3.51 (3.29)	1.86	2.05	1.67	1.07
	Control (n=14)	2.53 (2.27)	3.16 (2.62)	2.27	2.03	1.11	1.21
Three-month to six-month	**Full sample (n=23)**	0.67 (1.8)	0.43 (2.21)	0.32	0.15	0.37	0.2
	Intervention (n=14)	0.91 (1.93)	0.51 (2.50)	0.40	0.17	0.47	0.2
	Control (n=9)	0.29 (1.61)	0.31 (1.8)	0.15	0.13	0.18	0.17
Baseline to six-month	**Full sample (n=23)**	3.07 (2.37)	3.41 (2.94)	2.29	2.0	1.3	1.16
	Intervention (n=14)	3.57 (2.59)	3.82 (3.15)	2.48	2.09	1.38	1.21
	Control (n=9)	2.29 (1.85)	2.78 (2.63)	1.83	1.75	1.24	1.06

Abbreviations: SD, standard deviation; ES, effect size; SRM, standardised response mean; COPM, Canadian Occupational Performance Measure; Perf, performance; Satis, satisfaction

5.7 Discussion

This study assessed the responsiveness of the EQ-5D-5L and COPM within a sub-acute stroke population and has shown both outcome measures are highly responsive to change following inpatient stroke rehabilitation. However, a reduction in responsiveness was demonstrated in the EQ-5D-5L post-discharge with some values indicating a worsening in HRQOL for the total sample, and for the intervention and control groups. The COPM administered at baseline and three-month follow-up demonstrated high levels of responsiveness for performance and satisfaction, with moderate values found for performance and minimal values found for satisfaction at six-month follow-up. Unfortunately, external responsiveness was not assessed in this Chapter as a result of insubstantial correlations between the EQ-5D-5L and the external criterion.

5.7.1 EQ-5D-5L

Mean scores

The mean and estimated marginal mean for the index values reported at baseline and discharge respectively, were similar to the mean values reported by Lu *et al*. (2016), who assessed the responsiveness of the EQ-5D-3L within a sub-acute stroke population receiving inpatient rehabilitation. Mean index values at admission were 0.166 and 0.435 at discharge (Lu *et al*. 2016). Of note, where studies have evaluated the psychometric properties of the EQ-5D-5L in a stroke population, index values and EQ-5D VAS scores were notably higher than those reported here (Janssen *et al*. 2013; Golicki *et al*. 2015a; Golicki *et al*. 2015b; Chen *et al*. 2016).

Goliciki *et al*. (2015a) examined the responsiveness of the EQ-5D-5L in the acute phase of stroke, with individuals assessed one-week post-stroke and four months later. Golicki *et al*. (2015a) reported a mean index value of 0.577 and 0.691, and EQ-5D VAS score of 54.3 and 60.7 at baseline and follow-up respectively. These are notably higher than the means and marginal means reported for the current study at baseline and three-month follow-up. This could be the result of differences in level of independence between recruited samples, with participants demonstrating a greater level of independence in the study by Golicki *et al*. (2015a) (Barthel Index, mean=75). In comparison participants in the current study had a mean FIM total score of 64.35 at baseline, indicating a moderate level of disability (Inouye *et al*. 2001). As such this study provides novel insight into the responsiveness of the EQ-5D-5L when used with stroke survivors with a greater level of dependence in completion of ADLs.

In addition, the current study involved only participants whose upper limb rehabilitation formed the main focus of their occupational therapy treatment. As such the lower levels of HRQOL demonstrated could be reflective of the sample base, with lower levels of HRQOL associated with impaired upper limb function up to six months and one-year post-stroke (Nichols-Larsen *et al*. 2005; Franceschini *et al*. 2010). Future research could stratify participants according to level of upper limb impairment to discern variability in levels of HRQOL.

Responsiveness

The responsiveness of the EQ-5D-5L across several time intervals allowed consideration for the different degrees of change noted and what these changes might demonstrate. A substantial degree of change was seen between baseline and discharge indicating improvement in index values and the EQ-5D VAS.

Responsiveness was greatest between baseline and discharge, which was expected with greater gains in HRQOL typically experienced during inpatient rehabilitation (Hopman *et al*. 2003; Madden *et al*. 2006). This study was set within the initial stages of stroke rehabilitation, and as such pronounced changes were expected in line with spontaneous recovery and directed rehabilitation following stroke.

The subsequent decrease in EQ-5D-5L scores after hospital discharge was in stark contrast with the continued improvements in performance captured by the FIM between discharge and three-month follow-up, noted in Chapter 4. The resultant responsiveness values indicated a worsening in HRQOL with negative values found for this time period. Studies have found decreases in HRQOL in the months following hospital discharge. Using the 36-item Short Form (SF-36), a generic HRQOL outcome measure, Hopman *et al*. (2003) found statistically significant decreases in several domains of the SF-36 six months after hospital discharge following stroke, following noticeable gains during inpatient rehabilitation. Similarly, Suenkeler *et al*. (2002) reported reduced overall HRQOL in participants six to 12 months post-stroke, with significant decreases in the social functioning and physical functioning domains of the SF-36.

Discharge from the hospital marks a significant transitional period as stroke survivors are faced with the reality of their impairments and adjusted living conditions, which may account for the lower index values seen between discharge and three-month follow-up. This has also been identified in qualitative exploration of the experiences of stroke survivors, who have described the immediate period following discharge from hospital as the most difficult (Ch'ng *et al*. 2008). This result is particularly pertinent following a recent publication by the Stroke Association (2019) which reported that

over half of stroke survivors in Northern Ireland feel abandoned once they leave the hospital setting.

Additionally, the reduction in EQ-5D-5L scores and negative responsiveness values post-hospital discharge could be demonstrative of the limited responsiveness of the EQ-5D-5L in a stroke population. The reduced ability of the EQ-5D-5L to detect more subtle changes in a post-rehabilitation setting has been noted by Pickard *et al*. (2006). The use of a stroke-specific HRQOL outcome measure would provide additional items of relevance to stroke survivors and could be a useful adjunct in clinical trials alongside the EQ-5D-5L (Carod-Artal 2012).

Chen *et al*. (2016) assessed the responsiveness of the EQ-5D-5L before and after a three- to four-week intervention and found an ES of 0.40 and a SRM of 0.63 for the index value and an ES of 0.30 and a SRM of 0.34 for the EQ-5D VAS, lower than those found for the current study. However, the sample included both sub-acute and chronic stroke survivors (median time post stroke [interquartile range] = 19.7 [0.4-94]). The wide variability in time post-stroke likely impacted on responsiveness values, with Chen *et al*. (2016) reporting that the EQ-5D-5L was more responsive when used by individuals in the sub-acute phase of stroke with a lower level of functional ability prior to the delivery of the intervention.

The lower index values at baseline for the current study potentially contributed to the greater level of responsiveness seen, with participants demonstrating increased potential for change. Findings for this study are in line with Pickard *et al*. (2005) who reported that more extreme changes in health will lead to greater levels of responsiveness. Conversely, this leads to potential subtle changes in the sample

being overlooked by the EQ-5D-5L, as discussed above with lower levels of responsiveness found post-discharge.

The assessment of responsiveness of the EQ-5D-5L completed by Golicki *et al*. (2015a) used the Barthel Index and modified Rankin Scale as external criterions. A change of 9.25 or more on the Barthel Index indicated either improvement or deterioration, depending on the direction of change. Using the Barthel index as an external criterion for the EQ-5D-5L, responsiveness of the index value for the improved group (ES=0.71; SRM=0.86) was similar to those found for the current study across treatment groups between baseline and three-month follow-up. The EQ-5D VAS demonstrated a lower level of responsiveness in comparison (Golicki *et al*. 2015a), which was also similar to that found for the full sample and intervention group in the current study.

Comparisons across responsiveness studies are however limited due to variable time post-stroke and timing between follow-up periods. In addition, the application of different external anchors will impact responsiveness analyses. Golicki *et al*. (2015a) used stroke-specific (modified Rankin Scale) and generic (Barthel Index) assessments of functional ability as anchors, which led to differences across the reported responsiveness values. An additional external criterion used in a responsiveness analysis of the EQ-5D-5L included the VAS from the Stroke Impact Scale 3.0 to assess participant perceptions of recovery; correlations between the differences in scores on the EQ-5D-5L and the Stroke Impact Scale were used to denote responsiveness (Chen *et al*. 2016).

5.7.2 COPM

Prioritised goals

Occupational performance problems fell predominantly within the self-care domain, with the goal of washing and/or dressing most frequently prioritised. Prioritisation of difficulties within the self-care domain was similarly reported by individuals in the sub-acute phase of stroke in the study by Schiavi *et al*. (2018). However, there were no problems identified within the area of socialisation, with fewer issues identified within the leisure domain overall (Schiavi *et al*. 2018). Whilst occupational problems within the leisure domain were also identified to a lesser extent in the current study, where they were identified they predominantly related to difficulties socialising. Cultural differences could account for the variations in the type of issues identified, with the study by Schiavi *et al*. (2018) completed in Italy. In addition, the participants in the study by Schiavi *et al*. (2018) demonstrated severe impairments in completion of ADLs which may have impacted on the types of occupations prioritised. As a result, improvements in socialisation may not be an immediate priority with individuals restricted by the hospital environment and reduced independence in completing everyday activities.

Waddell *et al*. (2016) explored the goals prioritised by individuals with chronic hemiparesis following stroke, taking part in an upper limb intervention trial. Occupational problems identified by the COPM were categorised under the domains of ADLs, instrumental ADLs, work or leisure, using the American Occupational Therapy Association framework (American Occupational Therapy Association 2014). The most commonly reported problem under the ADLs domain was dressing, similar to the current study, with communication and home management the most frequently reported problems under instrumental ADLs. Although participants from both studies

were recruited to an upper limb effectiveness trial it is not clear to what degree this influenced responses. However, problems identified in the current study were activity- and/or participation-related with no upper limb impairment-level problems identified, indicating it may not have specifically influenced the problems identified.

Responsiveness

There has been limited evaluation of the responsiveness of the COPM within a stroke setting. A large degree of responsiveness was demonstrated between baseline and three-month follow-up, indicating improvement in occupational performance. Similar to the EQ-5D-5L and the activity-level outcome measures examined in Chapter 4, there was little change indicated between three- and six-month follow-up. Despite the minimal change between three- and six-month follow-up, scores on the COPM performance and satisfaction domains remained less than six indicating remaining difficulties in occupational performance. At this stage it is likely stroke survivors may have less input from community stroke teams and experience persistent difficulties in ADLs completion and engagement with roles in the community.

To our knowledge the responsiveness of the COPM in a stroke only population has not been examined. Multiple studies have examined mixed populations with many of these taking place in an outpatient setting or mixed inpatient and outpatient settings (Bodiam 1999; Wressle *et al*. 1999; Chen *et al*. 2002; Eyssen *et al*. 2011; Tuntland *et al*. 2016). Those studies which reported the numbers of stroke survivors recruited were generally of a low number (sample size <20) (Bodiam 1999; Chen *et al*. 2002; Tuntland *et al*. 2016).

An early validation study assessed the responsiveness of the COPM within a neurological inpatient setting with 17 participants, seven of whom were stroke survivors (Bodiam *et al*.1999). Using the paired t-test this study found there were significant differences in performance and satisfaction scores between admission and discharge. An assessment of the responsiveness of the COPM with a larger sample size (n=225) took place in a community setting (Tuntland *et al*. 2016). However, stroke survivors only accounted for 18 of the included participants (Tuntland *et al*. 2016). As a result, sample heterogeneity and differences in time post-stroke limits comparisons between these studies and the current study.

External responsiveness was assessed by Eyssen *et al*. (2011), referred to as 'criterion responsiveness' by study authors, across mixed outpatients pre-occupational therapy treatment and three months later. Although comparison with the current study is limited, a transition index was used to determine participant perceptions of change on the performance problems identified, indicating actual, meaningful change. Using an external criterion directly related to the construct under investigation, the COPM was able to distinguish between improved and unimproved participants.

The patient-reported rating of change administered during the pilot RCT was captured at one time point, prior to discharge, and related exclusively to recovery of upper limb function. As such it was not possible to examine external responsiveness for the COPM in the current study. Future studies should utilise both internal and external responsiveness methods in order to determine the ability of the COPM to capture meaningful change in a stroke population. An external anchor directly related to the COPM is recommended, such as the transition index used by Eyssen *et al*. (2011).

5.7.3 Limitations

Whilst both internal and external methods are recommended in the assessment of responsiveness (Husted *et al*. 2000), only internal responsiveness was assessed in the current study. Further research is required to determine the responsiveness of the EQ-5D-5L and COPM with the use of multiple anchors, incorporating both patient- and clinician-perceived ratings of change.

There are many factors linked to HRQOL following stroke and which could have potentially influenced scores on the EQ-5D-5L. Predictors of HRQOL include, but are not limited to: gender, comorbidity, mood disorders, fatigue, social support, shoulder pain, functional ability and upper limb impairment (Carod-Artal and Egido 2009). In the present context of an upper limb intervention pilot study it was not possible to explore all these factors in relation to HRQOL. However, consideration of these predictors in future study would be beneficial in determining treatment effectiveness of mirror therapy.

During collation of responses for the COPM not all occupational problems were assigned under a domain in the answer booklet, prior to listing problems on the scoring aspect of the booklet. As the PhD researcher, BT, did not collect the data, it was not possible to check with participants which domain they would individually assign these performance problems. Therefore, BT made a subjective decision, and assigned problems based on previous categorisations, which contravenes the client-centred focus of the COPM.

As reported in Chapter 4, attrition became a prominent problem when assessing responsiveness at each subsequent time point as the study progressed. By six-month

follow-up half the participants recruited to the control group remained, with 75% remaining in the intervention group. This potentially impacted on responsiveness values by six-month follow-up, with more pronounced differences seen in group responsiveness post-hospital discharge.

5.8 Conclusion

The responsiveness of the EQ-5D-5L and COPM was investigated in the sub-acute phase of stroke, and both outcome measures were considered suitable for use within three months of stroke onset. The goals identified by participants using the COPM were attributed low scores at three- and six-month follow-up, highlighting persistent difficulties in the areas of self-care and mobility. Taken alongside the reduction in HRQOL, there are potential implications for the level of rehabilitation and support provided to stroke survivors in the community.

This study found that the EQ-5D-5L was responsive to improvement in HRQOL during inpatient rehabilitation and the COPM was responsive to improvements in occupational performance up to three months follow-up. However, minimal change and responsiveness demonstrated by both outcome measures by six-month follow-up indicated their limited clinical utility in the community. External responsiveness was not completed, and as such the values found for the EQ-5D-5L and COPM should be used with caution.

This study provides valuable information regarding the delivery of PROMs for informing an RCT of mirror therapy following stroke. This study supports the administration of the EQ-5D-5L at baseline, discharge and three months, alongside a stroke-specific health-related quality of life outcome measure with consideration of the

reduced and negative responsiveness scores found on the EQ-5D-5L at three months. Additionally, this study supports the administration of the COPM at baseline and three-month follow-up, which could potentially benefit from fidelity checks to ensure recording of the COPM remains client-centred and in line with its theoretical base.

Chapter 6

Chapter 6 – Stroke survivors' views on mirror therapy in upper limb rehabilitation

6.1 Abstract

Introduction

Mirror therapy provides a novel, cost-effective approach to upper limb treatment following stroke. However, the effectiveness of mirror therapy in sub-acute stroke is unclear. Qualitative studies enable exploration and optimisation of intervention acceptability and implementation. This study explored the experiences of individuals who received mirror therapy as part of their upper limb rehabilitation.

Method

A qualitative approach was used to explore barriers and facilitators of mirror therapy. Three individuals were recruited from one hospital and took part in a focus group led by an experienced moderator. The participants were four-, 12- and 18-months post-stroke. The focus group was audio-recorded and transcribed verbatim. Thematic analysis was used to analyse the data.

Results

Three key themes were identified, namely: long-term recovery, healthcare barriers and practical considerations. Of note, participant experiences of mirror therapy were linked to their ongoing recovery, with physical and psychological factors impacting on intervention acceptability. Barriers to mirror therapy were posed by clinician decision-making with limited participant involvement, and lack of knowledge held in inpatient and community settings. Design aspects of the mirror box used impeded independent use by one participant with severe impairment. Further analysis of the themes posited

a systems-based framework to improve mirror therapy delivery and to support sustainable independent practice, based on education, collaborative goal-setting and co-production.

Conclusions

Mirror therapy was generally considered an acceptable upper limb treatment. However, participants held different viewpoints regarding when mirror therapy could best be applied, with acceptability negatively impacted by increased upper limb impairment and reduced sensation. A person-centred approach across healthcare settings is recommended to improve implementation and to facilitate self-management practices from hospital to home.

6.2 Chapter overview

This chapter was undertaken to explore the views of stroke survivors who had received mirror therapy as part of their inpatient rehabilitation. This study was designed to complement the studies which make up this thesis as part of the feasibility aims of a pilot RCT to examine the effectiveness of mirror therapy in the first three months of stroke. The researchers involved in this study are detailed in Appendix 1.

6.3 Introduction

The previous chapter assessed the responsiveness of PROMs and provided insight into the changes in HRQOL following inpatient rehabilitation and hospital discharge. In addition, occupations prioritised by individuals in the sub-acute phase of stroke were highlighted. The study demonstrated meaningful outcome measures were responsive and suitable for use in a main RCT. Although often overlooked (Price-Haywood *et al*. 2019), inclusion of PROMs in a stroke trial are integral when considering the effectiveness of an upper limb treatment, demonstrating aspects of patient-centred care (Duncan Millar *et al*. 2019). The Institute of Medicine (2001) described patient-centred care as “care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions.” Promoting the inclusion of patient preferences has also stimulated the exploration of recipient viewpoints when evaluating the effectiveness of an intervention (Sekhon *et al*. 2017). The Medical Research Council guidelines recommend the evaluation of acceptability during feasibility testing of complex interventions (Craig *et al*. 2008), with qualitative studies regularly used in the investigation of complex interventions (O’Cathain *et al*. 2015).

Exploring patient experience acts to unpick the context both internally and externally to an intervention and is an effective tool to identify ways to refine intervention design and procedure (Lewin *et al*. 2009; Hubbard *et al*. 2016). Giving a voice to the service-user is not only empowering but gives insights into user acceptability and their perceptions of clinician actions (Sekhon *et al*. 2017). An intervention must be acceptable to service-users for it to be relevant and contribute positively to treatment effectiveness (Yardley *et al*. 2015).

There has been limited research into the acceptability of mirror therapy. In a study by Horne *et al.* (2015a), the feasibility and acceptability of patient-led mirror therapy for the upper limb, and patient-led exercises for the lower limb, were explored in the acute phase of stroke, incorporating both quantitative and qualitative measures of acceptability. Patient-led therapy was used to describe exercises which were completed by stroke survivors in an in-patient setting without direct guidance or supervision from treating therapists. Stroke guidelines recommend a high dosage of therapy (Intercollegiate Stroke Working Party 2016), with at least 45 minutes per day across all relevant therapies recommended in UK stroke guidelines (Intercollegiate Stroke Working Party 2016). However, numerous studies report patients do not receive the recommended amount of therapy during inpatient rehabilitation (Åstrand *et al*. 2016; Clarke *et al*. 2015; West and Bernhardt 2012). Thus, the study by Horne *et al*. (2015a) aimed to increase the amount of time participants engaged in active therapy though the promotion of independent practice. In the study by Horne *et al*. (2015a) 94 participants, at least one-week post-stroke (range=7-33 days post-stroke), were randomised to receive either standard care plus mirror therapy or standard care plus lower-limb exercises. A semi-structured questionnaire was used to explore participant experiences of patient-led therapy following completion of the 28-day

treatment programme. Seventy-eight stroke survivors completed the questionnaire and of those randomised to receive mirror therapy, 70% (n=33) reported the exercises as useful and 86% (n=43) would recommend the treatment to other stroke survivors. Four weeks following the end of treatment, in-depth exploration of participant viewpoints (n=20) were captured using semi-structured telephone interviews. Although mirror therapy was not the sole focus of discussion, facilitators and barriers for mirror therapy were identified. Facilitators of mirror therapy were linked to a positive influence on patient autonomy. In addition, improvements in function increased participant motivation encouraging further engagement with mirror therapy. Barriers noted for mirror therapy were a lack of certainty regarding the use of visual illusion to improve motor function, and reduced motivation to engage, particularly when improvements in function were not seen. Practical barriers to using mirror therapy included difficulty setting up equipment independently and a lack of space to accommodate equipment.

The practical issues noted by stroke survivors in Horne *et al*. (2015a) such as lack of space to set up the mirror and difficulty managing items independently, such as exercise booklets, were also endorsed by staff (Horne *et al*. 2015b). Exploration of staff experiences implementing the patient-led therapy programme described above were explored by Horne *et al*. (2015b) using focus groups.

The study by Horne *et al*. (2015b) explored staff viewpoints on training participants to implement their individualised exercise programmes outside formal therapy sessions to encourage independent practice while in hospital. Minimal input was provided by healthcare staff several times per week through the provision of support and advice. Five focus groups were held across five locations once the programme had been

active for six months with a total of 30 participants. Although themes raised were in relation to delivering a patient-led programme, barriers for mirror therapy were noted. A large mirror set up in a large, quiet room was a strategy implemented by some staff members to help individuals complete the activities with less distraction. However, this strategy could only be taken up by more mobile stroke survivors.

Given the lack of evidence to underpin how mirror therapy can be delivered, and to whom, for optimum effectiveness (Thieme *et al*. 2018), qualitative exploration of patient acceptability of mirror therapy has the potential to inform how mirror therapy is implemented in clinical, home and research settings. Investigation of stroke survivors' experiences of mirror therapy is therefore required.

6.4 Aim and objectives

6.4.1 Aim

The aim of this study was to explore the views and experiences of patients who received mirror therapy as part of their in-patient stroke rehabilitation.

6.4.2 Objectives

The objectives of the study were:

1. To explore participants' experiences of mirror therapy as a component of upper limb rehabilitation using focus group methodology;
2. To evaluate participants' perceptions of the mirror therapy programme;
3. To examine the perceived barriers and facilitators influencing adherence to mirror therapy.

6.5 Method

6.5.1 Study design

This study used a descriptive qualitative methodology, commonly used for intervention development and refinement (Neergard *et al*. 2009). This was chosen due to the key focus being on exploring the views of stroke survivors on the phenomena under investigation, without the application of a conceptual theory or interpretative-heavy analysis (Sandelowski 2010). This process focused on the specific topic of mirror therapy and highlighted participant views on this treatment using a focus group method, which is recommended for exploratory research (Doody *et al*. 2013). Focus groups promote participant interaction and act as a means to highlight similarities and differences in opinions and experiences (Sim 1998). A focus group method enables participants to question each other, to build on answers and potentially lead to the formation of ideas and concepts which may not have been uncovered through other methods (Plummer-D'Amato 2008).

6.5.2 Positionality

Interpretation of the data remains an essential part of descriptive qualitative methodology (Sandelowski 2010), as one cannot truly separate the researcher from the study as the 'data collection instrument' (Bourke 2014). The use of reflexivity is important due to the potential impact of the researcher on the study design, implementation, analysis and write-up (Horsburgh 2003). Therefore, BT's positionality was described to promote transparency and allow readers to determine its impact on the study (Appendix 15).

6.5.3 Ethics and governance

The study procedure and protocol were approved by Ulster University Institute for Nursing and Health Research Filter Committee. Ethical approval was obtained from the Proportionate Review Sub-committee of the London - South East Research Ethics Committee on 21st November 2018 (Appendix 16). Governance approval was subsequently gained from the participating Health and Social Care Trust (Appendix 17)

6.5.4 Recruitment

Purposive sampling (Etikan *et al*. 2016) was used to recruit individuals with experience of the mirror therapy treatment, received as part of their inpatient rehabilitation in one hospital site in the NHSCT.

Any individual who had received mirror therapy for upper limb rehabilitation within the previous 12 months were eligible for recruitment. This was to ensure participants would be able to recall their inpatient stroke rehabilitation. See Appendix 18 for details of the intervention participants received.

The inclusion/exclusion criteria were as follows:

Inclusion criteria:

- Participant willing and able to give informed consent for participation in the study.
- Age ≥ 18 years.
- Diagnosis of stroke.
- Discharged from stroke rehabilitation ward in previous 12 months.

- Received mirror therapy as part of their upper limb treatment during in-patient stroke rehabilitation, during the sub-acute phase of stroke.

Exclusion criteria:

- Unable to follow and understand two step commands in English language.
- Substantial cognitive or communication difficulties.

Occupational therapists involved in the rehabilitation of stroke patients identified and screened potential participants from their caseload over the preceding 12 months. Invitation letters (Appendix 19) and participant information sheets (Appendix 20) were posted to those who met study criteria with a stamp addressed envelope for return of reply slip. Where there was no response from potential participants, reminder letters (Appendix 21) were sent two weeks following postage of the initial invitation letter. Those individuals who returned reply slips indicating interest in hearing further about the study, received a telephone call from BT and verbal consent was gained. Demographic information was collected from participants on the phone indicating their age, gender and date of stroke (Appendix 22).

Written informed consent (Appendix 23) was obtained from each participant prior to the commencement of the focus group.

6.5.5 Participants

Twenty individuals received invitation letters to take part in the study. The PhD researcher received eight responses from individuals who declined any further contact and five responses from individuals who consented for further contact to be made by the researcher. Following telephone contact with potential participants, three individuals were able to take part in the focus group. Whilst efforts were made to reach

a consensus for a focus group date, due to scheduling conflicts two individuals were unable to take part.

All three participants were male, aged 66, 77 and 58 years old; and were 12-, 18- and four-months post-stroke respectively.

6.5.6 Data collection

A topic guide was developed to explore participant views and experiences of mirror therapy, taking into account experiences of how it was implemented and perceptions of treatment effectiveness (Appendix 24). The topic guide was aimed at promoting discussion amongst focus group participants and questions were semi-structured and open-ended. Members of the academic and clinical research team, in conjunction with two service users who received mirror therapy as part of their upper limb stroke treatment within the NHSCT, were consulted to inform the topic guide. The topic guide acted as a dynamic document and included prompts. This was not prescriptively adhered to so as to allow for the exploration of additional meaningful conversations where they arose.

6.5.7 Setting

The focus group took place in a private room on the hospital site in the NHSCT, due to familiarity of the setting for all participants, with travel costs reimbursed.

The focus group was led by an experienced moderator, Leona Robinson (LR), who had not been involved in the treatment of the participants, who used the topic guide to guide the discussion. The PhD researcher was also present to take consent at the beginning and to act as an observer. The PhD researcher took reflective notes during

the session, including non-verbal communication and group interactions, to enhance the transcription and analytical process (Onwuegbuzie *et al*. 2009). This was augmented by discussion between LR and BT immediately after the focus group to discuss the proceedings and allowed LR to provide further feedback.

6.5.8 Procedure

Once written consent was obtained, LR and BT welcomed participants and completed introductions. Ground rules for the focus group were established before the group discussions began. Participants were reminded of the confidential nature of the focus group and were advised to respect the viewpoints of other group members. The focus group topic guide was used by LR to lead the discussion.

Informal member checking (Cohen and Crabtree 2008), which acts as a way to promote study transparency and rigour, took place once the focus group ended. For this LR summarised the key points raised and asked if the participants felt this was an accurate reflection of the discussion. The focus group lasted 90 minutes. Study rigor was enhanced through emphasis on participant viewpoints and interactions within qualitative description methodology (Milne and Oberle 2005).

6.5.9 Data analysis

The focus group was audio-recorded and transcribed verbatim by the PhD researcher. Guidelines for transcription were followed to aid data organisation and analysis (McLellan *et al*. 2003). These included: verbatim transcription, inclusion of pauses and all verbal, non-verbal and background sounds. All references to participant identities and hospital location were anonymised, to maintain confidentiality and anonymity. Transcription began on the day of the focus group to instigate understanding of what

was said. Transcription was completed by the PhD researcher to start the process of familiarity and immersion (Braun and Clarke 2006). The transcript was checked by a member of the research team who was not present for the focus group for accuracy to ensure study rigor.

Thematic analysis was used to analyse the transcript, following the guidelines by Braun and Clarke (2006). The data was coded manually with thematic analysis led by BT. This began with familiarisation during the process of transcription, which involved repeatedly listening to the audio recording and reading the transcript. There was review of the transcript, during which annotations were made in the margins and sections of the data were coded (Spencer *et al*. 2014). The list of codes was collated into categories, and subsequently assigned themes. These themes were continually reviewed, until nothing new was gained.

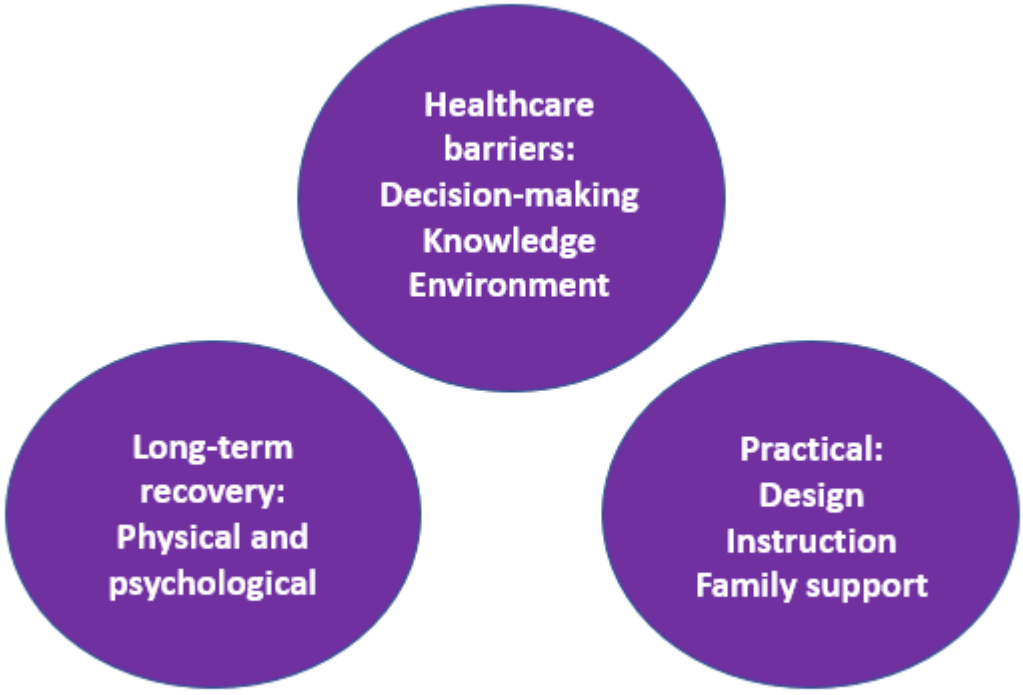
Supplemental to the thematic analysis completed by BT, LR independently completed thematic analysis of the transcript. This involved reading the transcript several times, coding the transcript and assigning themes. Once completed the two researchers (BT, LR) met to discuss themes and reached consensus on those identified. An extract of the transcript from this study has been included in Appendix 25.

6.6 Results

Participants viewed mirror therapy as a beneficial treatment modality, the themes were predominantly linked to their broader experiences of upper limb rehabilitation and their journey of recovery following stroke. All themes were underpinned by autonomy and active participation. The following three themes were identified from the data collected:

long-term process of recovery; healthcare barriers and practical considerations (Figure 6.1).

Figure 6.1 Core themes identified from the thematic analysis



6.6.1 Long-term process of recovery

During the focus group discussion, mirror therapy was linked to participants' experiences of their long-term recovery with upper limb recovery described as an ongoing process. This process was influenced by physical and psychological factors, which acted as either facilitators or barriers of mirror therapy. Emblematic of the heterogeneity of stroke, participants had differential viewpoints on how these factors impacted on their experiences of mirror therapy. Physical factors included the participants' level of functional ability and continuing changes in ability over time.

Psychological factors included how participants' experiences of mirror therapy were impacted by motivation, fatigue and additional support.

6.6.1.1 Physical factors

Participants were introduced to mirror therapy during the sub-acute phase of stroke. At this early stage post-stroke, the upper limb may not be able to functionally contribute to rehabilitation activities. Two participants reported how delivery of mirror therapy helped facilitate autonomy as no minimum level of functional ability was required to take part. For some participants mirror therapy provided increased control and engagement, positively driving upper limb recovery.

> *"…mentally I felt I still was active, it was just my left side I couldn't use and that fitted in very well with what was being done through the mirror box."* (P2)

However, one participant reported a discord in terms of the upper limb activities on offer and their level of ability. The participant perceived that reduced sensation in their upper limb impacted on their ability to successfully engage with mirror therapy and felt mirror therapy would have been more beneficial if received at a later date. In this situation the participant acknowledged the benefit of receiving targeted therapies at different stages of their recovery.

> *"I think I would have liked… sort of thing now, now I'm getting some sensations down my arm. It, it might be more beneficial at this stage, but certainly while I was in hospital, I was getting very little… reaction at all out of my fingers and my hand."* (P1)

Participants acknowledged the long-term nature of their recovery, as they continued to experience changes in function. However, for one participant they perceived a lack of recognition from inpatient and community therapists of the potential for improvement in their upper limb. The perceived reduced focus on their upper limb rehabilitation while in hospital continued once they were discharged home. This was further demonstrated by the fact one participant was not provided with a mirror box for use at home. The remaining two participants were provided with a mirror box for use in hospital and to take home. However, overall participants felt there was a lack of acknowledgement of the benefit of targeted therapies at different times, in line with individual goals and level of impairment.

6.6.1.2 Psychological factors

Engagement in occupation was integral to participants' sense of autonomy, through both upper limb rehabilitation and carrying out day to day activities. During inpatient rehabilitation, participants reported mirror therapy provided them with the opportunity to be active. There was a strong sense of the importance of individually driving their rehabilitation gains, with their efforts being imperative for improvement in upper limb function, which was facilitated through mirror therapy.

> *"...gave me a focus because it was all about rehab you know and unless I was making an effort to get myself better, it wasn't going to happen automatic so I was pleased I was doing something."* (P1)

Autonomy was also reflected beyond the therapy setting. Some participants reported mirror therapy encouraged movement in the more affected upper limb, and later facilitated transfer-effects, encouraging use of the limb in ADLs. The use of mirror

therapy assisted some participants in reaching the long-term goals of using their upper limb in ADLs.

> *"…and I couldn't use my left hand but what I found was I was attempting to replicate what I saw the right hand doing…"* (P2)

> *"And it- it was the prompt I needed to get me started to thinking about what I should be doing."* (P2)

In contrast, for another participant autonomy through rehabilitation was better achieved through exercise-based activities, which involved direct movement of the more affected upper limb. Participants described an understanding of the theoretical background of mirror therapy and the use of visual feedback to promote improvement in upper limb function. However, this appeared to conflict with their views of the recovery process.

> *"The other thing that helped me was I had like a skateboard thing; you could move around the table. I got more satisfaction out of that because I could feel my arm moving… that was very positive… I suppose from like being younger you always felt like if you hurt your leg or your something like that, it was the physical exercise that got you going…."* (P1)

Participants predominantly received therapist-led inpatient rehabilitation, with most reporting reduced motivation to complete mirror therapy activities independently at home. One participant noted a preference for direct therapist support and guidance completing rehabilitation activities, both in hospital and at home.

> *"I find it so much easier whenever the stroke team of the OTs and the physios coming to the house to do the exercises… much easier than trying to do it on my own on a daily basis, because it it's very easy to turn round and say look, look I'll leave it to tomorrow."* (P3)

The support of family members was noted as important to assist with completion of mirror therapy exercises while in hospital. In addition, the importance of the continuation of upper limb rehabilitation activities once discharged home was explicitly stated by all participants. For the participants there was no end date for their recovery and was an ongoing process requiring endurance.

> *"I think there is still an awful lot of hard work has to go in… and I think that's the hardest part… since coming out of hospital it's being able to set aside um your hour or couple of half hours in the day to go and do exercises with… the weak side"* (P3)

However, there was no reference to family involvement in upper limb activities once participants were at home. This highlights a potential gap in therapists working with stroke survivors and family members in the development of long-term treatment.

The ongoing process of recovery included navigating rehabilitation alongside fatigue. For one participant they reported feeling a sense of guilt when they were unable to engage in activities independently following outpatient appointments.

> *"… I was in for three full days of um intensive physiotherapy… I didn't do anything at home… and come Tuesday I felt guilty then and so you're actually*

then trying to force yourself to go back to doing exercises and maybe pushing yourself too much at times and doing too much." (P3)

Supportive mechanisms were valued by participants in order to establish autonomy over, and maintain engagement with, their upper limb rehabilitation. Cognisant of the supportive impact family and therapists provide for stroke survivors, it was noteworthy for participants when they experienced less supportive input from healthcare staff.

6.6.2 Healthcare barriers

Participants spoke about occupational therapists providing access to upper limb treatments and rated positively the opportunity to engage in a range of treatments, including mirror therapy. However, some participants reported therapist decision-making could act as a barrier to upper limb rehabilitation, such as allocated time spent on upper limb activities and treatment modalities chosen.

For one participant there was an imbalance between the amount of time spent on their upper and lower limb treatments, at the expense of their upper limb. While walking was a primary goal for increased autonomy, improvements in their upper limb function remained important to the participant and required additional therapist input.

"I would have liked to have more time spent on my arm. Because I still have very little movement, movement in it. But for the walking it was great. And sorta once I could do one thing… what's next… But I felt I needed more exercises on my arm, there wasn't enough of that." (P1)

Participants highlighted mirror therapy was undermined by a lack of knowledge about the treatment modality amongst the wider hospital staff. While participants reported

the negative perceptions did not affect their uptake of the therapy, this could have potential implications for others.

> *"Unfortunately, there is a scepticism out there on the ward about the mirror box, too many people didn't know the thinking behind it and I thought that was a shame."* (P2)

> *"I think if, the OTs understood it ok, it was really the other care and nursing staff didn't really get the hold of it."* (P1)

Participants also reported a lack of knowledge about mirror therapy amongst therapists in the community. Provided with a mirror box for use at home, a participant reported community therapists did not see the benefit of mirror therapy and as such he felt discouraged from using it. Experiencing increased pain and reduced range of motion at his shoulder following hospital discharge, the participant felt that lack of encouragement to use mirror therapy, and additional tasks which involved gross motor activities, contributed in part to his current state.

> *"I would have to say that the stroke community team… I don't know whether they knew about the mirror box or what, but they thought that I was well past that and I should be moving onto doing more difficult exercises… I think that the OTs and the physios… kind of dis-encouraged me from using it..."* (P3)

> *"I know I was able to do things before I left hospital, and I can't unfortunately do those now…I, I honestly think that if I was to have kept the mirror box on*

and uh the other things that I was doing, the movement in the arms um, hand etc., that I wouldn't be in the position that I am now." (P3)

The actions of healthcare staff were additionally reported as impacting on individual autonomy. The placement of a participant's bed against a wall restricted their ability to use their more affected limb in ADLs while in hospital. For the participant, encouragement and opportunity to make therapeutic gains was not for therapy sessions alone but something to be extended throughout the hospital environment.

"You talked about your- your left-hand side what they did with me was they pushed me into a corner, but the bed was long ways on and my left-hand side was against the wall. So yeah, I mean I didn't know. I- I knew my left-hand side wasn't working you know; I didn't need to be saved from it." (P2)

6.6.3 Practical considerations

Some discussion questions focused on the practical aspects of mirror therapy, including the design of the device, the instruction manual provided, and exercises completed. All participants agreed that the manual provided was useful and the exercises were simple, with some reporting they were able to memorise the movements. Participants were predominantly content with the overall size of the mirror box, which could be folded and stored in their bedside lockers, promoting ease of use and storage.

"The fact it folds down so neatly you were able to keep it up the locker beside your bed." (P2)

However, another participant found the mirror box was not able to support their hemiplegic limb and required additional therapist support when in use. The mirror box is made of polyester, a lightweight material. Consequently, one participant required therapist support to maintain their upper limb in a fixed position inside the mirror box. Without this additional support the participant's arm would move off the edge of the table. The ease of use noted as a benefit by some participants, acted as a barrier with an individual with a greater level of upper limb impairment.

> *"I found it hard to do it my own because whenever I put my left arm in that thing it, it would automatically fall off the table and tip the mirror box over."* (P3)

To overcome practical barriers, suggestions were made to use a non-slip mat to secure the mirror box in place, with varying experiences of utility amongst participants. Increasing the surface area of the mirror itself was also suggested, to improve focus on the mirror image.

Participants all acknowledged family members assisted with upper limb activities while in hospital. However, concerns were raised about family members not understanding the premise of mirror therapy and preferring to complete physical activities.

> *…the family found it easier to do the skateboard thing with me cause… see it more as physical. Then once they get over the initial shock… can I do this, can I do this well it was all physical stuff.* (P1)

6.7 Contextual analysis

Although the aim of this study was to complete a descriptive qualitative analysis of a single intervention, the findings indicated a need for further exploration of the barriers and facilitators which impacted on mirror therapy acceptability.

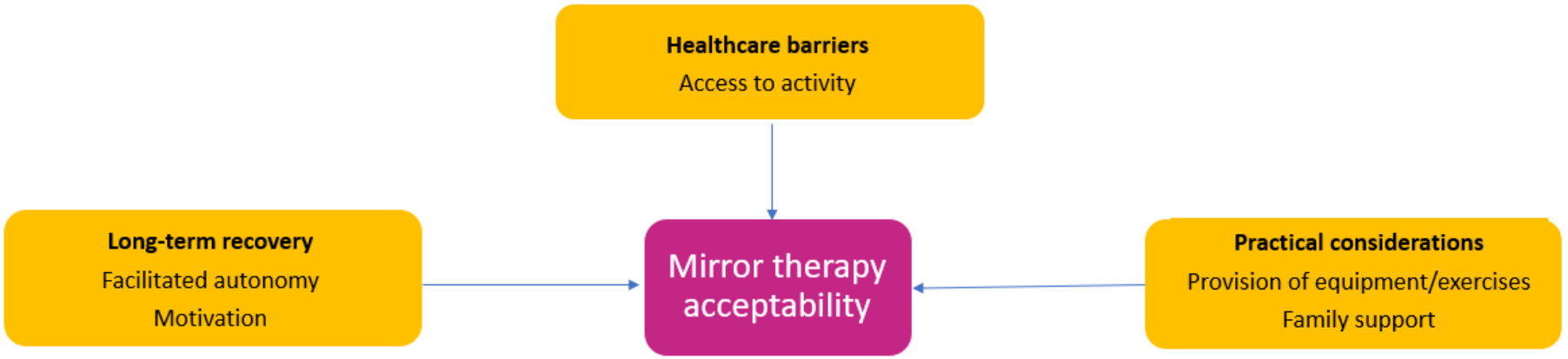
Participants demonstrated an openness to engage with mirror therapy and provision of the mirror box device and exercise booklet gave most of the participants the impetus to use the treatment outside of formal therapy hours and with family support. Facilitators and barriers for the acceptability of mirror therapy often overlapped (Table 6.1), which highlighted the complexity of upper limb recovery following stroke.

Table 6.1 Themes summarising the barriers and facilitators of mirror therapy

Themes		**Barriers**	**Facilitators**
Long-term process of recovery	**Physical**	Upper limb impairment Changes in function	Upper limb impairment Changes in function
	Psychological	Fatigue Guilt What comprises upper limb treatment?	Afforded autonomy Motivation Family support
Healthcare barriers		Lack of knowledge across settings Inadequate communication across settings Therapist decision-making	Therapist decision-making
Practical considerations		Size of mirror Lightweight design Family support	Provision of mirror box Exercise booklet Lightweight design Family support

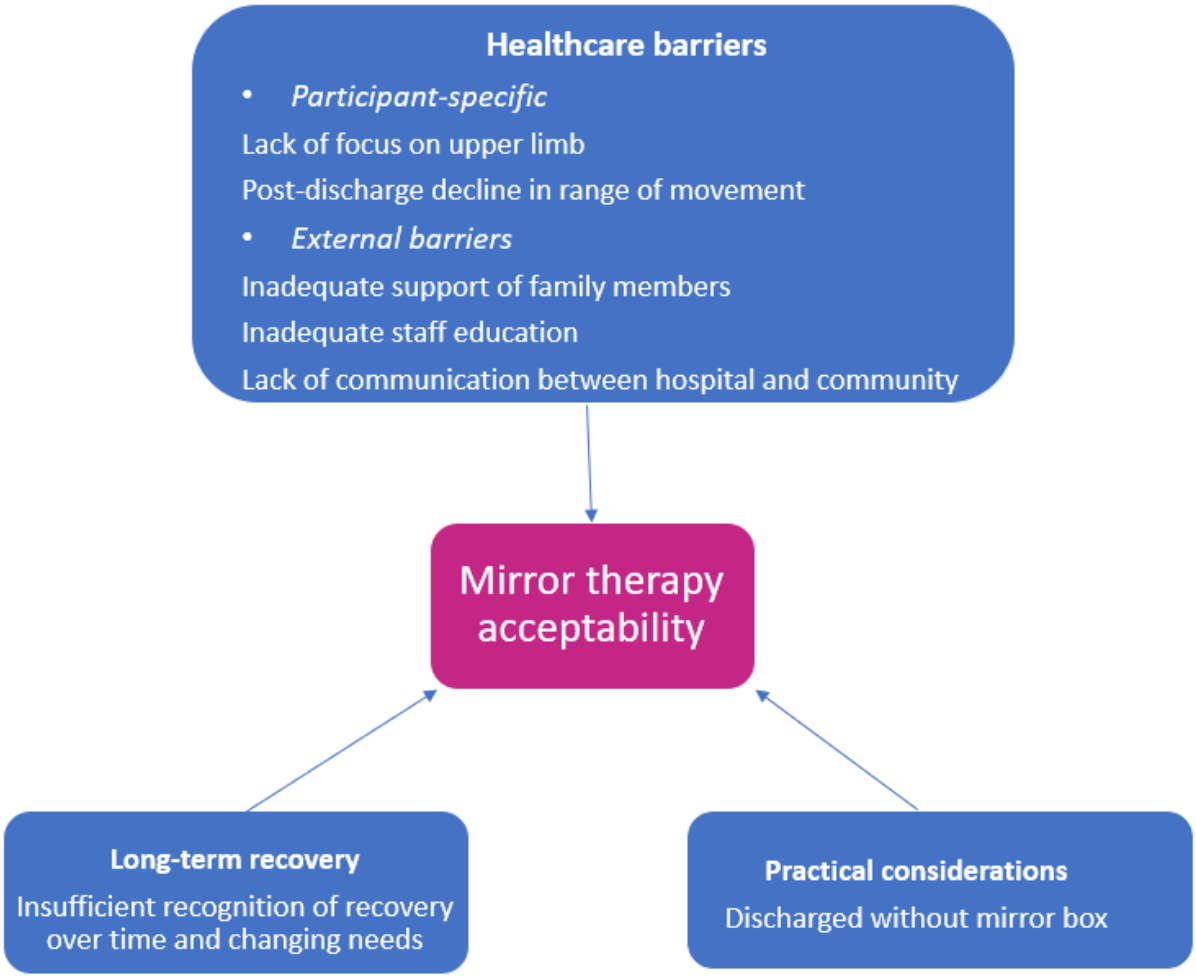
Although difficulties were reported with therapist decision-making, it is important to acknowledge that participants also spoke positively of their experiences with clinicians. Participants recognised the role clinicians played in providing access and support to complete upper limb rehabilitation activities and in fostering autonomy. As such it is vital to build on these relationships and encourage collaboration between stroke survivors, family members and the stroke clinical care team, and prioritise of individualised treatments. An overview of the facilitators of mirror therapy are shown in Figure 6.2.

Figure 6.2 Facilitators of mirror therapy acceptability



The thematic analysis of the focus group demonstrated limitations in the delivery of integrated stroke care across the recovery continuum for participants. Figure 6.3 illustrates the contextual issues which impacted on mirror therapy acceptability for each core theme. Integrated patient care has been defined as “patient care that is coordinated across professionals, facilities, and support systems; continuous over time and between visits; tailored to the patients’ needs and preferences; and based on shared responsibility between patient and caregivers for optimizing health” (Singer *et al*. 2011). Although Northern Ireland has been at the forefront of integrated healthcare in the UK, it is not clear how patient-centred the delivery of that care is (Ham *et al*. 2013). Wide funding disparities have been identified across health and social care programmes in Northern Ireland with priorities heavily influenced by the medical model culminating in acute services receiving the most funding, (Ham *et al*. 2013). The Regulation and Quality Improvement Authority (RQIA) review of stroke services in Northern Ireland highlighted inadequacies and lack of co-ordination in the care patients received across Trusts (RQIA 2014). Some of the issues included: lack of communication and co-ordination between primary and secondary care; limited access to self-management programmes; limited communication with, and involvement of, family members; and lack of therapy provisions in the longer-term (RQIA 2014). These issues mirrored the barriers participants reported in the current study and emphasised the need for improvement in the overarching healthcare system. The findings from the current study indicated there was inadequate interpersonal integration, which is integration among health professionals, patients and family members within and across different settings, and inadequate process integration, referring to the extent to which patient care was coordinated across settings and over time (Singer *et al*. 2020).

Figure 6.3 Contextual factors which impacted on mirror therapy acceptability



When considering integrated care, it is important to note the complexity of the healthcare system and therefore the inherent complexity of delivering a single healthcare intervention within it. The WHO described six building blocks through which the relationships and interactions among them make up a healthcare system: service delivery; health workforce; information; medical products, vaccines and technologies; financing; and leadership and governance (de Savigny and Adam 2009). As a result, multilevel approaches have been recommended to improve healthcare delivery, acting through the different levels of the healthcare system (Ferlie and Shortell 2001). One such approach designed to improve the integrated care of individuals with chronic disease, is the chronic care model (Wagner 1998). The chronic care model identified six essential components for optimal healthcare: self-management support, decision support, delivery system design, clinical information systems, community resources and health care organisation (Bodenheimer *et al*. 2002). Systematic reviews have found improvements in health outcomes and healthcare practice following implementation of components of the chronic care model across chronic conditions (Davy *et al*. 2015; Reynolds *et al*. 2018).

It is beyond the remit of this PhD to develop a chronic care model-based intervention for stroke survivors with upper limb impairment. However, the model provides a useful lens to examine what improvements could be made to mirror therapy implementation based on the core themes. The components of self-management support, decision support and delivery system design seemed most pertinent to the findings of the current study. This is largely due to the focus on the education and support of stroke survivors and their family members to manage their condition in the long term; the use of evidence-based clinical guidelines to guide care for an informed clinician; and a

structure which enables multidisciplinary team members to work collaboratively with stroke survivors and their family members with regular follow-up.

Deeper reflection identified education, collaborative goal setting and co-production as factors across the core themes to improve the delivery of mirror therapy central to person-centred practice.

Education

A lack of education and knowledge of mirror therapy held by healthcare staff was highlighted by participants, which can lead to difficulties in knowing when to implement mirror therapy and with whom. A lack of understanding of mirror therapy outside the inpatient stroke rehabilitation unit was also highlighted.

> *"I think the OTs that used the mirror box were great. But, unfortunately there is a scepticism out there on the ward about the mirror box, too many people didn't know the thinking behind it and I thought that was a shame."* (P2)

> *"I would have to say that the stroke community team… I don't know whether they knew about the mirror box… but they thought that I was well past that and I should be moving onto doing more difficult exercises."* (P3)

Training on mirror therapy at ward level would be recommended across the multi-disciplinary team and supporting staff. In addition, the extension of training is recommended across different occupational therapy departments to ensure effective carryover of mirror therapy regimes from inpatient to outpatient settings.

Related to education of staff is the importance of educating patients and family members. While individuals will have differential experiences of mirror therapy, as demonstrated by opposing participant views on when best to receive mirror therapy, ineffective communication should not be a contributing factor.

Education of family members should also be implemented to assuage any misconceptions regarding mirror therapy. One participant reported family members preferred physical activities over mirror therapy.

> *"…I think they… the family found it easier to do the skateboard thing with me cause I think family sort of see it more as physical."* (P1, when discussing activities completed with family members in hospital)

Furthermore, there was no mention of family members assisting with mirror therapy once participants were discharged. This demonstrated the potential gap in family members being informed and involved in the upper limb rehabilitation of stroke survivors to prepare for the increase in self-directed therapy upon hospital discharge.

Collaborative goal setting

Collaborative goal setting offers a vital path for stroke survivors to make meaningful contributions to their stroke care. Participants reported limited involvement in the development of their treatment plans, exemplified by inadequate communication with healthcare staff in hospital and community settings. This led to participants feeling insufficient treatment options were provided to them which led one participant to report there was not enough attention paid to their upper limb, and another to link this to greater degrees of upper limb impairment.

"I would have to say that the stroke community team had said to me… about the mirror box… they thought that I was well past that… I honestly think that if I was to have kept the mirror box on and uh the other things that I was doing, the movement in the arms um, hand etc., that I wouldn't be in the position that I am now" (P3)

"… I would have liked… more exercises for my arm …overall of which the mirror box should be a part of it." (P1)

Collaborative goal setting acts as a means to take on board the views of stroke survivors, to identify their needs and work through solutions in partnership. Participants voiced differing viewpoints with regards to when mirror therapy can best be applied. With improved training and education, therapists can have open and informed conversations with patients and family members to enable them to make informed decisions about their rehabilitation. Therapists working collaboratively with stroke survivors and family members has the potential to positively impact their experiences of mirror therapy and influence treatment acceptability.

Co-production

Co-production illustrates a further element of person-centred practice, which can be used to improve the implementation of mirror therapy and upper limb stroke services. Through co-production, stroke survivors and family members are viewed as equal partners in the design and delivery of treatment. Participants reported therapists as determining which upper limb treatments to offer and when, with participants, although motivated, relegated to a more passive role in their rehabilitation.

Linked to this, participants reported feeling less engaged in completing mirror therapy activities independently, with one participant reporting a preference for therapist-directed upper limb treatments, both in hospital and once home.

> *"…I found it much, much easier whenever and even since I got out… of hospital, I find it so much easier whenever the stroke team of the OTs and the physios coming to the house to do the exercises and put me through my paces, um I find that much, much easier than trying to do it on my own on a daily basis…"* (P3)

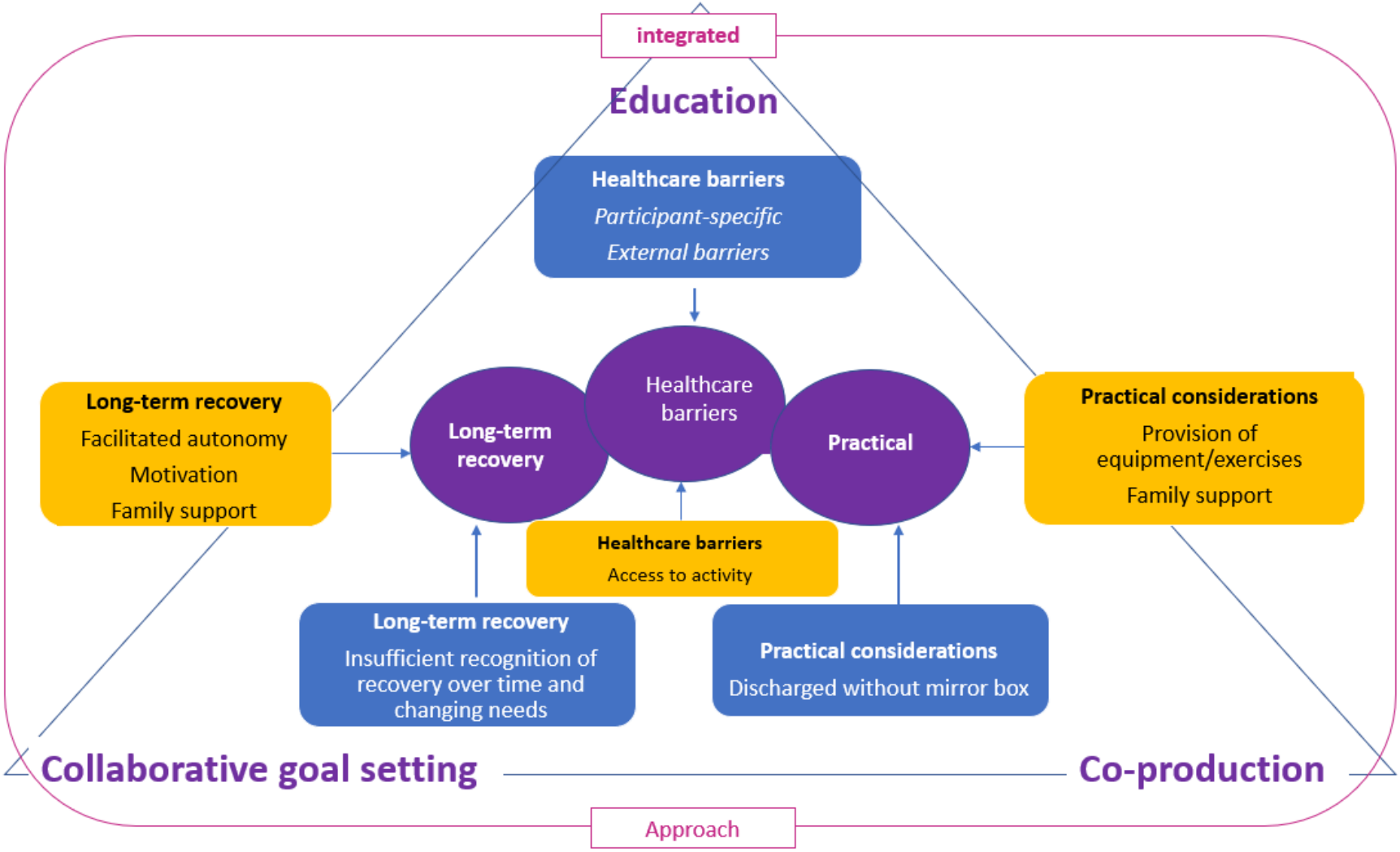
In recognition of the long-term impact of stroke, rehabilitation must allow for active participation and involvement from those directly affected to encourage and facilitate greater investment in their own care and empowerment. Furthermore, the expert experiences of stroke survivors can be used to design a mirror therapy box which requires minimal therapist support. One participant was unable to implement mirror therapy independently, which posed a barrier while in hospital and has far reaching implications for an individual's ability to implement self-directed practice.

Participants reported negative remarks about mirror therapy and an environment not conducive to independent practice of rehabilitation activities. Co-production would aim to promote a supportive rehabilitation environment, to allow stroke survivors to engage in rehabilitation activities outside of formal therapy sessions and encourage individual engagement in meaningful day to day activities.

6.7.1 Framework for mirror therapy implementation

Drawing on the findings from the core themes and the facilitators and issues highlighted above, a framework was proposed to improve the implementation of mirror therapy as part of upper limb rehabilitation (Figure 6.4). Education, collaborative goal setting and co-production were hypothesised as essential elements to optimise delivery and uptake of mirror therapy as part of this integrative framework (Figure 6.4). Drawing on components of the chronic care model, stroke survivors and their family members would be supported by knowledgeable healthcare professionals across settings, and rehabilitation and self-management programmes would be delivered in line with evidence-based guidelines for improved transition into the community and beyond (Bodenheimer *et al*. 2002). Illustrative of the chronic care model, this framework would require changes in how stroke care is delivered to support stroke survivors and their family members manage their upper limb treatment in the long term. Future research could use this framework to design a mirror therapy intervention incorporating changes in how the intervention is developed and designed across the stroke care pathway.

Figure 6.4 Developing an integrated approach framework for implementation of mirror therapy



6.8 Discussion

The aim of this study was to explore stroke survivors' experiences of mirror therapy as part of their upper limb rehabilitation and examine the facilitators and barriers that may impact on its acceptability. This study indicated that while mirror therapy was considered a useful adjunct to upper limb rehabilitation, barriers to its acceptability were posed by inadequate knowledge held by healthcare staff and lack of consideration for the individual needs and preferences of the participants. Furthermore, participants often reflected on their overall experiences of upper limb therapy and the continuing implications of stroke. The themes identified were long-term impact of stroke, healthcare barriers and practical considerations. Building on the themes identified, a framework for the successful implementation of mirror therapy was devised. Incorporating an integrated approach, changes in how stroke services are delivered would involve key consideration for education, collaborative goal setting and co-production.

As noted in Chapter five the transition home following stroke can be especially fraught, as stroke survivors adjust to changes in their life situation (Ch'Ng *et al*. 2008). Beyond impacting on a person's functional ability, stroke survivors experience challenges to their identity and roles, requiring adaptation and adjustment (Salter *et al*. 2008; Hole *et al*. 2014). Pallesen (2014) described the trajectory of stroke as being one of continual change and as such participant reference to their current state of being could be considered unsurprising. Participants in the focus group described difficulties adjusting to the increased responsibility held over their upper limb treatment and navigating conflicting feelings of guilt for not doing enough and fatigue once at home.

All participants reported ongoing changes in their upper limb function and suggested their recovery was a continually evolving process, in line with Pallesen (2014).

Participants stressed the need for continued rehabilitation through self-led activity and, where possible, active use of the more affected upper limb in daily activities. Referencing recovery as a continual process requiring sustained rehabilitation activity was similar to the findings reported by Barker and Brauer (2005). Barker and Brauer (2005) described stroke survivors as working towards recovery across the lifespan. Participants remained hopeful that with sustained activity and support, their upper limb function would continue to improve. The study by Barker and Brauer (2005) took place more than 10 years ago, however similar issues with the restricted time frame through which stroke recovery services are delivered continue to be identified (Teasell *et al*. 2012; Luker *et al*. 2015).

To support participants with self-led activity in hospital and at home, clinicians gave the mirror box device to some of the study participants. However, participants reported a reduced likelihood to engage with mirror therapy outside formal therapy hours and once at home. Reductions in adherence to a patient-led mirror therapy programme were found for individuals in the acute phase of stroke, where they perceived no improvement in function (Horne *et al*. 2015a). Reduced compliance to therapeutic interventions has also been found among stroke survivors in the community once the programme ended (Touillet *et al*. 2010; Jurkiewicz *et al*. 2011; Miller *et al*. 2017). Multiple factors influence adherence to home-based therapy programmes, including social support, self-efficacy, motivation, and the patient-clinician relationship (Scorrano *et al*. 2018; Essery *et al*. 2017). In light of the healthcare barriers experienced by participants in the current study, the theme for which is explored

below, future study could assess how these factors impact on long-term adherence to mirror therapy. Beyond provision of a mirror box to two of the participants, it is not clear how participants and their family members were prepared in hospital for independent delivery of mirror therapy once at home.

The support of family members in continuing mirror therapy and additional upper limb activities during inpatient rehabilitation was viewed as crucial. The provision of a manual for the exercise regimen enabled individuals to engage in mirror therapy outside of formal therapy hours and with family members. However, notably missing from the focus group discussion was family involvement in upper limb activities once participants were home, and explicit discussion preparing participants and family members for the increased responsibility over their upper limb treatment. Studies have highlighted the integral contribution family members provide in increasing the activity levels of stroke survivors while in hospital (Maeshima *et al*. 2003; Harris *et al*. 2010; Galvin *et al*. 2011). In the study by Galvin *et al*. (2011), an additional exercise programme was delivered by family members, alongside standard treatment for the duration of inpatient rehabilitation. The intervention group demonstrated improvements in lower limb range of movement and walking ability, and reduced caregiver burden. The benefits of involving family members during inpatient rehabilitation to support the long-term management of stroke were further demonstrated by Kalra *et al*. (2004). Caregivers were trained in moving and handling techniques and how to facilitate completion of ADLs during inpatient stroke rehabilitation, with improved psychosocial outcomes demonstrated for both caregivers and stroke survivors up to one-year post-stroke (Kalra *et al*. 2004).

Family members play an important role in the recovery of stroke survivors; providing support to complete ADLs and to cope with the psychosocial impact of stroke

(Moreland *et al*. 2009; Satink *et al*. 2015). Increased levels of social support have been linked to greater and faster degrees of functional improvement following stroke (Glass and Maddox 1992; Ng *et al*. 2013), and further work could explore how family members can support stroke survivors to engage with self-led mirror therapy outside formal therapy sessions and in the community.

Participants in the current study identified the importance of rehabilitation as an occupation and the autonomy mirror therapy afforded them in assuming responsibility over their recovery. This finding was similar to the study by Purcell *et al*. (2018) where rehabilitation was viewed as an important occupation, vital for recovery and something which stroke survivors wanted to actively engage with.

Proot *et al*. (2007) explored stroke survivors views of how clinicians facilitated their autonomy at different stages of stroke recovery, from the acute setting through to discharge. Proot *et al*. (2007) described rehabilitation as being when stroke survivors begin to increase their autonomy and play a more active role in this process. With this in mind, mirror therapy provided a suitable means for participants to increase autonomy and control. A systematic review of 31 qualitative studies explored stroke survivors' experiences of inpatient physical rehabilitation (Luker *et al*. 2015) and found similar themes to the current focus group study, with physical activity and autonomy particularly valued (Luker *et al*. 2015).

Participants were explicit in the current study of the personal role they held in driving their recovery. Maclean *et al*. (2000) explored the attitudes of 22 stroke survivors undergoing rehabilitation who exhibited either high or low motivation. Maclean *et al*. (2000) suggested stroke survivors who were highly motivated were more likely to view

rehabilitation as the most important factor for their recovery and perceive an active role in this process.

It is also worth noting all participants reported scepticism towards mirror therapy when it was first introduced, and it was only later when improvements were noted that they felt the therapy was of benefit. Horne *et al*. (2015a) also noted participants were motivated to engage further with mirror therapy following improvement in function. As such, a greater degree of upper limb impairment and slower response to treatment could pose a barrier to its uptake.

In addition, a barrier may be posed by the premise of mirror therapy, which is based on visual illusion. A participant in the focus group who had experienced a greater degree of upper limb impairment remained sceptical of mirror therapy throughout their rehabilitation. This participant reported a preference for activities which involved direct movement of the more affected upper limb and reported the same of their family members. This highlights the need for clinicians to effectively communicate the theory behind mirror therapy to ensure understanding for both stroke survivors and family members. This is of particular importance as belief in the visual illusion could play a role in the effectiveness of mirror therapy (McCabe 2011). Due to the non-traditional aspect of treatment, McCabe *et al*. (2011) described how mirror therapy could be introduced to individuals with chronic regional pain syndrome alongside introductory exercises, to encourage familiarisation. Future research could adapt these strategies for stroke survivors and family members to ease initial scepticism.

Participants in the current study identified the importance of support from healthcare staff for upper limb rehabilitation. However, the actions of staff were mostly notable

through the barriers they posed. While a participant perceived adequate focus on their lower limb rehabilitation, this was in stark contrast to a lack of attention paid to their upper limb. This was exemplified further when they were not provided with a mirror box for use at home. Similar qualitative findings were reported by Barker and Brauer (2005) where stroke survivors reported reduced focus on upper limb treatment in comparison to the lower limb, with reduced recognition for the potential of upper limb recovery. A recent qualitative study exploring the viewpoints of stroke survivors and healthcare professionals regarding stroke upper limb rehabilitation also suggested there was a disproportionate focus on the lower limb and linked this to the pressures to discharge patients as soon as possible (Meadmore *et al*. 2019).

A barrier to mirror therapy was posed by a lack of effective communication and person-centred practice by healthcare staff. Participants reported negative remarks were made by healthcare staff about mirror therapy in hospital and community settings. Luker *et al*. (2015) noted the disempowering impact poor communication can have on stroke survivors, and Maclean *et al*. (2000) suggested conflicting messages may adversely impact motivation to engage with rehabilitation. Health professionals play an important role in treatment acceptability (Sekhon *et al*. 2017) and in the recovery journey of stroke survivors (Proot *et al*. 2000). As such healthcare staff should be aware of the impact their behaviour can have on patient motivation and treatment adherence.

Healthcare professionals acting as gatekeepers, determined which upper limb treatments stroke survivors had access to and when, which posed an additional barrier. A participant in the current study felt they had received mirror therapy at too early a stage in their recovery and could potentially benefit from use of the treatment

at the present time. Another participant reported they were dissuaded from using mirror therapy in the community by healthcare professionals who cited the participant's current level of function as too advanced for the treatment. This study highlighted the lack of knowledge that remains regarding the level of upper limb impairment to which mirror therapy can best be applied, and at which stage following stroke (Thieme *et al*. 2018). Furthermore, mirror therapy may not be suitable for all and this study emphasises increased choice and involvement for stroke survivors in the development of their upper limb treatment plan.

This study was unable to assess the factors involved in the deterioration of upper limb function following hospital discharge experienced by one participant. However, the lack of collaboration in goal setting was reported as a potential factor by the participant. Stroke guidelines for the UK advocate for person-centred goal planning, with links to increased motivation, self-efficacy and greater engagement and satisfaction with the rehabilitation process (Rosewilliam *et al*. 2011; Sugavanam *et al*. 2013). In the study by Luker *et al*. (2015) person-centred goal planning was an important theme linked to patient motivation and autonomy, thereby stressing the importance of collaboration for encouraging self-management among stroke survivors.

The final theme identified related to the practical aspects of delivering mirror therapy. Participants demonstrated acceptability for the exercises that were delivered due to their simplicity, promoting memorisation and noted the benefits of an exercise manual. A recommendation was made to increase the size of the mirror to cover all sides of its supporting structure to improve concentration on the visual illusion. A recent secondary analysis of a meta-analysis of 32 stroke trials assessed which aspects of mirror therapy implementation influenced its effectiveness (Morkisch *et al*. 2019). Morkisch *et al*. (2019) found that a larger mirror, unilateral exercises and those which

do not involve object manipulation were linked to increased effectiveness. This has implications for future development of a mirror box device.

The mirror box is foldable and lightweight, which while recognised as a facilitator by some participants, became a barrier for another, as the box would slide off the table during exercises. Recommendations were made by a participant to use a non-slip mat to provide increased support. However, there were indications the mirror box provided may not be suitable for independent practice by individuals with a greater degree of upper limb impairment, which was also identified by Horne *et al*. (2015a). In the UK increased emphasis has been placed on service user involvement to improve health care services, with co-production forming an element of this (Crawford *et al*. 2002; Needham and Carr 2009). Building on patient-centred care, future study could draw on elements of co-production and co-design to formulate a mirror box device that meets the needs of stroke survivors with severe upper limb impairment (Nasr *et al*. 2015). Nasr *et al*. (2015) demonstrated how the perspectives of stroke survivors can be used in the design of robotic technologies for stroke survivors through a participatory approach, which involved consideration for user requirements, preferences and goals.

The barriers found for mirror therapy suggests an integrated approach across health professionals and settings is needed to better prepare stroke survivors and family members for long-term recovery and to provide them with the information needed to successfully engage with mirror therapy and self-manage their treatment. Additional analysis of the data identified three factors which could improve mirror therapy delivery across inpatient and community settings: education, collaborative goal setting and co-production. The principles of the chronic care model informed the framework designed

to improve mirror therapy implementation, centring the patient throughout for improved transition into the community with the necessary support and resources to self-manage their treatment and recovery.

The chronic care model was designed to improve the management of chronic conditions involving integration across providers, the community and the healthcare system (Bodenheimer *et al*. 2002). Studies have shown improvements in patient outcomes and levels of care following a chronic care model-based intervention with various chronic conditions (Coleman *et al*. 2009; Hopman *et al*. 2016). The implementation of the model's components aims to enable patients to actively engage in their healthcare, equipped with the knowledge and skills to better manage their condition and for improved interactions with expert healthcare providers (Bodenheimer *et al*. 2002).

A narrative review found that the implementation of interventions underpinned by components of the chronic care model has the potential to ease the transition between hospital and ambulatory services for older individuals (Sendall *et al*. 2017). Aspects of the interventions included collaboration between community and allied health services, community resources, evidence-based practice, patient education and self-management (Sendall *et al*. 2017). Future study would be needed to evaluate the effectiveness of the proposed framework to support mirror therapy implementation to ease the transition from inpatient rehabilitation to community care.

6.8.1 Limitations

Due to the late development of this study following conclusion of the pilot RCT, it was not possible to consider the acceptability of mirror therapy in relation to the feasibility

testing of the pilot study. As such, a broader analysis of acceptability incorporating both quantitative and qualitative methods was not possible. Of those recruited to the current study one participant had taken part in the pilot RCT. As a result, potential differences may have occurred in the dosage of mirror therapy received by participants with trial monitoring procedures not in effect. A qualitative study embedded in a main RCT of mirror therapy would be advised to build on the findings of the current study, to ensure standardisation of treatment received and to increase the sample base.

This study aimed to recruit 24 individuals and complete three focus groups with eight participants in each group. However, potential participant numbers were over-estimated, with a final number of 20 individuals who were considered eligible for recruitment. All participants received written study information by post followed by a reminder letter; the use of telephone to contact non-responders could be implemented to improve future recruitment rates (Nystuen and Hagen 2004; Harris *et al*. 2008).

While one focus group was completed with three participants, this accounted for a relatively high recruitment rate of 15%. However, with an all-male, small sample group, there is the potential additional themes could have been identified with a larger, more heterogenous sample. In addition, the expansion of the mirror therapy protocol across additional hospital sites in a larger trial would enable exploration across a more diverse sample base.

Following closure of the focus group study it emerged that a participant was recruited who did not meet the inclusion criteria and who had received inpatient rehabilitation more than 12 months prior to recruitment. This situation was discussed with the sponsor of the study, Ulster University Research Governance. The sponsor advised

this element of criteria was not related to harm reduction, but rather to ensure participants would be able to recall and participate fully in the focus group.

Eligibility was confirmed by clinical members of the research team who completed screening procedures at one time-point. Newington and Metcalfe (2014) completed a thematic meta-synthesis of the perceptions of researchers and clinicians to the recruitment of participants and highlighted the importance of engendering a research community. This could involve increased face to face contact between clinicians and researchers ensuring ongoing training and awareness of recruitment procedures. Fletcher *et al*. (2012) reported clinician understanding of research may pose problems which could be improved through increased training. In addition, the role of clinicians in research may be seen as an added burden to their daily roles (Newington and Metcalfe 2014). A practical way to reduce any burden could involve the easy access to eligibility criteria such as use of a visual aid on the wall of the department, avoiding the navigation of paper-based study materials. In addition, evaluations of eligibility criteria could be completed at multiple time-points, including during initial telephone contact by the researcher.

A write-up of the incident is included herein and is included in the study master file. Furthermore, a decision was made to retain the participant's data in the analysis because informed consent was provided, and the participant was able to contribute fully to all aspects of the focus group discussion.

6.9 Conclusion

This study contributes to the dearth of research concerning acceptability of mirror therapy. This is of relevance due to uncertainty regarding its effectiveness (Thieme *et*

al. 2018). Mirror therapy was identified as an acceptable treatment overall when delivered alongside a range of upper limb treatments. Autonomy and being active underpinned all identified themes. Emphasis was placed throughout on stroke as a chronic condition, requiring ongoing management and the need for support from healthcare professionals to engage in rehabilitation and ADLs. Although primarily delivered as a supervised treatment, there was scope for the participants to self-direct mirror therapy outside of formal therapy hours while in hospital and in the community. The current study identified the need for a supportive rehabilitation environment, from improved communication and collaboration between healthcare staff and patients, through to the physical layout of the ward itself to facilitate independent practice and autonomy.

Rehabilitation is aimed at improving a stroke survivors' physical, cognitive and psychosocial functioning for optimum participation and quality of life (NICE 2013; Intercollegiate Stroke Working Party 2016). As such it is important future research considers how mirror therapy can best be delivered to support patient autonomy and activity. In addition, as stroke survivors spend less time in hospital it is imperative a person-centred approach is used to help stroke survivors prepare for the transition home with increased responsibility over their long-term recovery.

Chapter 7

Chapter 7 - Discussion

7.1 Introduction

This PhD was undertaken to support the objectives of a pilot RCT examining the effectiveness of mirror therapy in the sub-acute phase of stroke. These objectives were related to the measurement properties of outcome measures included in the trial and, following completion of the pilot trial, examination of the acceptability of mirror therapy among stroke survivors. The systematic review completed in Chapter 2 highlighted the need to investigate the psychometric properties of the gWMFT, an outcome measure used in the pilot trial to assess upper limb function. The gWMFT demonstrated acceptable reliability, with increased measurement error found for assessing performance time between raters and at two different time points (Chapter 3). The responsiveness of all outcome measures included in the pilot trial; the FIM, gWMFT, EQ-5D-5L and COPM, were investigated to assess their ability to capture change across the assessment time points included in the pilot trial and demonstrated acceptable responsiveness (Chapters 4 and 5). Finally, the acceptability of the mirror therapy protocol delivered in the pilot trial was investigated and highlighted the need to inform staff members on the ward and in the community of mirror therapy to support delivery of the trial and potentially enhance adherence to the treatment (Chapter 6). This chapter aims to incorporate and discuss the key findings from each of the studies completed as part of this PhD thesis and discuss the implications for clinical practice and future research.

7.2 Summary of main findings

Chapter 1

With no high-quality evidence for any upper limb treatment for stroke (Pollock *et al*. 2014), mirror therapy is a unique intervention to potentially improve upper limb function in individuals with a range of abilities using visual feedback (Thieme *et al*. 2012). A Cochrane review demonstrated mirror therapy may be effective in improving upper limb function (Thieme *et al*. 2018), however heterogeneity across included studies meant firm conclusions could not be made about its effectiveness. With few studies examining the effectiveness of mirror therapy in the sub-acute phase of stroke, pilot trials are essential in building the evidence base for mirror therapy and supporting the development of robust clinical trials.

Chapter 2

A systematic literature review was completed and published which examined how the graded Wolf Motor Function Test (gWMFT) has been used and included examination of its psychometric properties (Turtle *et al*. 2019). The review identified 12 studies, all of which involved stroke survivors. Reliability was the only psychometric property assessed in two studies (Bonifer *et al*. 2005; Pereira *et al*. 2015).

The review highlighted the gWMFT is a complex tool involving multiple test items. However, most of the studies failed to adequately report essential elements of test administration and scoring, thereby hindering comparison of transparent and meaningful results. This review highlighted the need to explore the psychometric properties of the gWMFT in a sub-acute stroke cohort to support its continued use in a main RCT.

Chapter 3

Examination of the reliability and agreement of the gWMFT was completed during active recruitment for the pilot RCT and has been published (Turtle *et al*. 2020). This

study demonstrated the gWMFT could be reliably scored through direct observation and video recording. However, raters for both inter- and intra-rater analyses had difficulty distinguishing between level A and level B test items, with inadequate agreement found for performance time. These differences indicated scoring accuracy may be negatively impacted by the implementation of a large-scale trial with multiple raters.

Chapter 4 and Chapter 5

Chapters 4 and 5 examined responsiveness, as defined by Husted (2001), involving the assessment of internal and external responsiveness. Chapter 4 investigated the responsiveness of the activity-level outcome measures: the FIM and the gWMFT. Chapter 5 investigated the responsiveness of the PROMs: the EQ-5D-5L and COPM.

Chapter 4

Ceiling effects were found for the FIM cognitive sub-scale at all assessment time points and demonstrated negligible responsiveness, indicating potential difficulties in accurately portraying cognition following stroke. The FIM and gWMFT demonstrated internal responsiveness, with minimal change observed post three-month follow-up.

The gWMFT FAS correlated with the patient-reported rating of change to denote external responsiveness and was able to discriminate between participants who had experienced important change and those who had not.

Chapter 5

Due to inadequate correlations with the patient-reported rating of change external responsiveness could not be examined. The EQ-5D-5L was most responsive to changes in health-related quality of life (HRQOL) between baseline and discharge

(SRM>0.8). The decline in scores for the EQ-5D-5L between discharge and three-month follow-up could be the result of minimal change occurring, with greater levels of responsiveness predominantly found when there are extreme changes in health (Pickard *et al*. 2005). However, this could also indicate a decline in HRQOL, with the transition home following stroke reported as the most difficult (Ch'ng *et al*. 2008).
The COPM reported participant goals in the sub-acute stage of stroke and were predominantly related to improvement in self-care tasks. Greater levels of internal responsiveness were found between baseline and three-month follow-up (SRM>1).

Chapter 6

This study explored stroke survivors' views on mirror therapy and was the first known study to examine perceptions of mirror therapy treatment during inpatient rehabilitation. Thematic analysis identified three themes: long-term impact of stroke, healthcare barriers and practical considerations.

In order to improve the implementation of mirror therapy, a framework was developed based on participant experiences. Education, collaborative goal-setting and co-production were central tenets of this framework. This framework has similarities to the chronic care model with the aim to better facilitate and support self-management practices, incorporating all relevant stakeholders (Bodenheimer *et al*. 2002).

7.3 Recommendations and implications for future research

There was a marked decline in the number of participants who remained in the pilot RCT by six-month follow-up. Although Chapters 4 and 5 were focused on psychometric assessment, future study could examine reasons for attrition in detail. Due to little change in scores occurring between three- and six-month follow-up across

all outcome measures, with results likely impacted by attrition, follow-up assessment at three-month only is recommended in a future trial. All outcome measures demonstrated adequate internal responsiveness for use in a main RCT for up to three months post-stroke.

High levels of inter-rater reliability for the gWMFT demonstrated two therapists could score the gWMFT with a similar ranking of scores using video or through direct observation. A standardised training programme for the implementation and scoring of the gWMFT would be recommended due to the disagreements which occurred at item level between raters, and to a lesser degree, by one rater scoring at different time-points.

Due to the substantial floor effects found for gWMFT performance time scores, an alternative method of demonstrating an accurate overview of performance time would be recommended. Hodics *et al*. (2012) proposed reporting performance rate, indicating the number of times a participant can complete a test item over one minute, with zero reported for individuals unable to complete the item. Future studies should consider using performance rate (Hodics *et al*. 2012), which has subsequently been used in investigations of the WMFT and the grade 4/5 version of the WMFT (Taub *et al*. 2013; Uswatte *et al*. 2018).

The FIM cognitive sub-scale may not be sensitive to post-stroke cognitive impairment, as evidenced by ceiling effects and negligible responsiveness. Belief in the visual illusion of mirror therapy may play a role in its effectiveness (McCabe 2011), and as such participants must demonstrate the cognitive capacity necessary to engage with treatment. Additionally, the prevalence of cognitive impairment is high following stroke

(Tatemichi *et al*. 1994; Jokinen *et al*. 2015) and an additional screening tool would be recommended in a main RCT to ascertain the presence of cognitive deficits.

The wider implication of the results reported in Chapter 5 is that outcomes prioritised by stroke survivors, such as the COPM, can be successfully implemented in a clinical trial. However, considering the decline in HRQOL by three-month follow-up the addition of a stroke specific HRQOL outcome measure is recommended. This could provide further insight into the particular difficulties stroke survivors face once home (de Wit *et al* 2015). Further to the reductions in HRQOL, quantitative and qualitative exploration of potential factors impacting on HRQOL could also be explored.

The use of a global rating scale for patients and clinicians to rate changes in health is recommended for future examination of external responsiveness across all outcome measures. The global rating of change question should be directly related to the outcome measure of interest and include a 7- to 11-point scale to illustrate the magnitude of change, as recommended by Kamper *et al*. (2009).

To build on the exploratory findings reported in Chapter 6 a process evaluation implemented as part of a large scale RCT would be recommended. Process evaluations allow for the examination of contextual factors impacting on treatment effectiveness and can include qualitative exploration of participant viewpoints as well as of those implementing the intervention (Moore *et al*. 2015). A process evaluation will enable detailed consideration for how mirror therapy is implemented across sites and to examine potential factors impacting on effectiveness (Moore *et al*. 2015).

The exploration of how clinicians can support and facilitate self-management practices from an early stage in rehabilitation is recommended. Further research could also develop the framework with a larger sample size incorporating the viewpoints of all relevant stakeholders and examine the structural changes necessary to better support self-management practices. In Northern Ireland a recent report highlighted 45% of all stroke survivors feel abandoned once they leave the hospital (Stroke Association 2019). Further work is therefore necessary to support stroke survivors and their caregivers as they navigate this difficult transition and self-manage their upper limb treatment.

7.4 Strengths

A strength of this thesis is the use of a range of research methods as part of the feasibility testing of mirror therapy for upper limb rehabilitation in the sub-acute stage of stroke and has gone on to inform the development of a main RCT (see revised protocol for main RCT in Appendix 26). The systematic literature review provided an in-depth overview of the gWMFT and was published by the British Journal of Occupational Therapy (Turtle *et al*. 2019). This review highlighted the context where the gWMFT could be considered appropriate for use and is of direct relevance to researchers and clinicians. Although there was a lack of research evaluating the psychometric properties of the gWMFT (Turtle *et al*. 2019), this study indicated where further research is necessary.

This PhD is the first to assess the reliability and responsiveness of the gWMFT within three months of stroke onset. The psychometric evaluation of all outcome measures was conducted rigorously, with detailed descriptions of the procedures followed. Furthermore, while the use of PROMs in stroke research remains small (Price-

Haywood *et al*. 2019), the evaluation of the responsiveness of the EQ-5D-5L and COPM in Chapter 5 provides evidence to support their use in research and clinical practice.

Although there were limitations to the acceptability study of mirror therapy, as discussed below, the inclusion of this study added a qualitative element to this PhD. The exploratory study offered direct insights into the experiences of stroke survivors who received mirror therapy as part of their upper limb rehabilitation and provided direction for further areas of inquiry. In addition, this was the first study to primarily explore acceptability of mirror therapy.

This PhD was embedded in a pilot RCT and BT joined an occupational therapy-led research team consisting of university researchers and clinicians. The PhD researcher maintained close links with occupational therapists at the research site at all stages of the project, including discussion of study development and presentation of study findings, which may act to strengthen research engagement among clinicians (Paget *et al*. 2017).

7.5 Limitations

Sample size

The cohort recruited to the pilot study numbered 40 participants, thus limiting the sample size available for the psychometric studies completed as part of this thesis. Furthermore, the reliability study (Chapter 3) was completed during recruitment for the pilot RCT, and thus was limited to 30 participants.

In Chapter 6 inclusion criteria for the exploration of stroke survivors' views on mirror therapy was limited to individuals who had received mirror therapy within the previous 12 months at one hospital site. As a result, 20 individuals met the inclusion criteria. Although, the recruitment rate was 15%, overall only three participants took part in the focus group study.

Level of upper limb impairment

As part of a pragmatic pilot trial, recruitment was open to all individuals who required occupational therapy treatment for upper limb function. The participants recruited to the study demonstrated a range of upper limb abilities and as such it was not clear to which cohort the gWMFT can best be applied. Additionally, this may have impacted on the floor effects found for the gWMFT performance time. The criteria for classifying upper limb impairment reported by Uswatte and Taub (2013) and Uswatte *et al*. (2018) could be implemented as part of the inclusion criteria in a future RCT, with the gWMFT suitable for use with individuals classified at grades 3/4.

Video recording quality

In Chapter 3 video quality was poor leading to the inability to score single test items and the removal of one participant from intra-rater analyses. This has the potential to impact scoring participants by video in future studies. It is recommended to upload all video files immediately and review all videos to check for visual clarity and ensure the video is complete.

External responsiveness

How responsiveness was assessed in Chapters 4 and 5 was limited as most participants were recruited prior to study development and additional measures could not be included. Therefore, external responsiveness was only assessed for the gWMFT functional ability scale. The use of multiple anchors incorporating patient-based, clinician-based and clinically relevant anchors would be recommended in future study, that would be able to illustrate meaningful changes in scores on the FIM, gWMFT, EQ-5D-5L and COPM (Revicki *et al*. 2008).

Violation of protocol

A participant was recruited to the focus group study who had received mirror therapy beyond the 12-month cut off point and thus did not meet inclusion criteria. This error was noted following study closure and sponsors of the study based at Ulster University were notified. Although, this aspect of the inclusion criteria was not related to harm reduction and no further action was to be taken, this was a violation of the study protocol approved by the Research Ethics Committee.

7.6 Impact

The findings of this PhD have fed directly into the development of a multi-site RCT investigating the effectiveness of mirror therapy in a sub-acute stroke population and are summarised in Table 7.1 (see Appendix 26 for main RCT protocol). Inclusion criteria now includes a cognitive screening tool and participants are required to demonstrate a minimum level of upper limb function. The National Institute of Health Stroke Scale is recorded by therapists to denote severity of stroke, which is recommended as part of the baseline measures recorded in early stroke trials (Kwakkel *et al*. 2017).

Due to the minimal changes in scores and level of attrition which occurred by six-month follow-up in the pilot RCT (Chapter 4 and Chapter 5), assessment at six-month follow-up was removed. Outcome measures are now assessed at baseline, every two weeks until discharge and at three-month follow-up.

The gWMFT is currently delivered by two occupational therapist researchers. To overcome scoring disagreements between raters as highlighted in Chapter 3, training videos have been developed with two service users through patient and public involvement in research, to ensure standardised implementation and scoring. Furthermore, the researchers meet regularly to review participant performances to date and discuss scoring, reaching a consensus when necessary.

The findings in Chapter 5, which illustrated a reduction in HRQOL, and in Chapter 6, where difficulties using mirror therapy post-discharge were highlighted, have informed the development of a funded PhD opportunity within Ulster University based around the involvement of caregivers in mirror therapy delivery in the community.

This research has wide implications for the delivery of mirror therapy in research and clinical practice, where there is no agreed method for delivery and the conditions for optimum effectiveness remains unclear.

Table 7.1 Summary of PhD findings and links to main trial

PhD findings	**Implications for main trial**
The gWMFT is a complex tool to administer (Chapter 2).	Standardised training provided throughout implementation of main trial, to support implementation and scoring of gWMFT.

The gWMFT demonstrated good inter- and intra-rater reliability (Chapter 3). Inadequate agreement found for scoring gWMFT performance time (Chapter 3).	Training videos to support implementation and scoring of gWMFT have been developed with service users. These will be used by researchers completing outcome measure assessments within the main trial. Researchers will meet regularly to discuss gWMFT scoring and review participant videos where necessary.
The gWMFT demonstrated acceptability reliability (Chapter 3). The FIM, gWMFT, EQ-5D-5L and COPM demonstrated acceptable responsiveness (Chapter 4 and Chapter 5).	The FIM, gWMFT, EQ-5D-5L and COPM are appropriate for use in a main trial.
Floor effects found for the gWMFT (Chapter 3 and Chapter 4).	Minimum level of upper limb function is required to accurately determine effectiveness of mirror therapy, and enable accurate assessment using the gWMFT. This will be incorporated into the main trial. Mean performance rate calculated for gWMFT performance time, as devised by Hodics *et al*. (2012) will be used in the main trial.
Floor effects and negligible responsiveness found for FIM cognitive sub-scale (Chapter 4).	Addition of cognitive screening tool as part of inclusion criteria of the main trial to support delivery of protocol.

Increased attrition as study progressed (Chapter 4 and Chapter 5).	Removal of six-month follow-up assessment in the main trial.
Negative staff attitudes towards mirror therapy (Chapter 6).	Mirror therapy in-service training will be provided to ward staff.

7.7 Implications for clinical practice

Psychometric evaluation of the outcome measures included in the pilot trial, provide important information for development of a main RCT. However, they also provide relevant information for clinicians and add to the evidence base for upper limb assessment tools following stroke. As an activity-level outcome measure, the gWMFT has the potential to provide information regarding a stroke survivors ability to use their affected limb in everyday tasks, with direct relevance for occupational therapists.

The gWMFT demonstrated acceptable reliability when used with individuals in the sub-acute phase of stroke and when scored by two therapists. A barrier for the gWMFT is posed by the requirement to video record patient performances for therapists to view to later score. in Chapter 3 the potential to score functional ability without the requirement of video recording participant performances was demonstrated which would increase the clinical utility of the gWMFT.

The FIM, EQ-5D-5L and COPM all demonstrated internal responsiveness, supporting their use as evaluative outcome measures in stroke rehabilitation. The gWMFT demonstrated internal and external responsiveness indicating it was able to capture meaningful changes in upper limb function and could be used to track patient progress during inpatient stroke rehabilitation.

The gWMFT is a complex tool to implement and score, as illustrated in Chapters 2 and 3. The fidelity check developed by Morris *et al*. (2009) for the WMFT could be adapted for the gWMFT to ensure standardised implementation. The checklist developed by Morris *et al*. (2009) consisted of sequential steps to follow for each item to minimise variation in how the WMFT is implemented, detailing verbal instructions, item demonstration and patient set-up. Furthermore, the differences in agreement between raters in Chapter 3 highlighted that standardised training for scoring the gWMFT may reduce agreements between therapists. The functional ability scale could undergo revision to improve the clarity of scoring criteria to minimise potential confusion differentiating between level A and level B items.

Studies have previously identified the benefits of the COPM in supporting client-centred practice and goal setting (Colquhoun *et al*. 2010; Donnelly *et al*. 2017; Enemark Larsen *et al*. 2019). This is of particular relevance for clinicians where inadequate communication and involvement in their upper limb treatment plans were identified as barriers to engaging with mirror therapy (Chapter 6). The psychometric support provided for the COPM in stroke rehabilitation (Chapter 5), further supports the use of the COPM as a client-centred, goal-setting tool to promote collaboration between patients and therapists.

Mirror therapy was viewed as a useful upper limb treatment (Chapter 6). However, participants expressed differing views of its effectiveness and when it can best be implemented. Patient and family education are important to accurately convey the possible mechanism for mirror therapy as well as the latest research about its effectiveness, to potentially improve acceptability and reduce scepticism.

Mirror therapy was reported as predominantly clinician-led and participants reported a lack of motivation to continue with their therapies independently following hospital discharge (Chapter 6). Therefore, it is necessary for clinicians to better prepare patients for discharge and integrate self-management techniques at an early stage in rehabilitation. Findings from Chapter 6 resonated with studies reporting the unmet needs of stroke survivors in the community (Pindus *et al*. 2018; Abrahamson and Wilson 2019). Individualised, patient-centred care was recommended, with therapists delivering care across the recovery spectrum in line with the changing needs of stroke survivors.

7.8 Conclusion

This thesis presents a body of research which built upon the feasibility aims of a pilot RCT examining the effectiveness of mirror therapy, and additionally explored the acceptability of mirror therapy among stroke survivors. The literature review identified how the gWMFT has been used in stroke trials and the psychometric evidence available to support its use. The quality of studies included were predominantly of a low quality and supported further research to examine the psychometric properties of the gWMFT before progressing to a main RCT.

Challenges are posed by the requirement to score participants completing the gWMFT using video. The inter-rater reliability study demonstrated a similar ranking of scores between raters and therefore scoring by video may not be necessary. However, inter- and intra-rater agreement analyses highlighted rater difficulties differentiating between level A and level B items. Therefore, standardised training delivered throughout a main

RCT would be recommended, potentially with a quality control process to reduce errors.

Responsiveness is an essential psychometric property for evaluative outcome measures and was examined in all outcome measures delivered in the pilot RCT. The FIM, gWMFT, EQ-5D-5L and COPM all demonstrated internal responsiveness and captured change across the time points under investigation. The use of an external criterion enabled exploration of external responsiveness in order to determine whether the change which occurred was meaningful. The gWMFT functional ability scale was the only outcome measure which correlated with the external criterion and demonstrated external responsiveness. Future investigation of responsiveness should include a range of criterions in order to investigate the ability of the outcome measures to capture clinically relevant change. Caution should be applied to responsiveness values by six-month follow-up due to the level of attrition, with recommendations for omission of six-month follow-up in a main trial.

In addition, the reduction in HRQOL between discharge and three-month-follow-up was notable and could be potentially reflective of the transition from hospital to home. Further research is required to examine the factors which may have impacted on the reductions in HRQOL.

The findings of this thesis suggest mirror therapy is an acceptable upper limb treatment. However, scepticism regarding the effectiveness of mirror therapy could impact on adherence and active involvement in therapy. Improved communication and collaboration across healthcare hierarchies and between stroke survivors and family members is recommended to support stroke survivors in their recovery.

This thesis supports the use of the gWMFT, FIM, EQ-5D-5L and COPM in a main RCT examining the effectiveness of mirror therapy in the sub-acute stage of stroke. Furthermore, this work has contributed to the evidence base for trial development of upper limb interventions and potentially influence how mirror therapy is delivered and assessed in practice.

References

Abrahamson, V. and Wilson, P.M. (2019) How unmet are unmet needs post-stroke? A policy analysis of the six-month review. *BMC Health Services Research*, *19*(1), 480.

Almhdawi, K.A., Mathiowetz, V.G., White, M. and delMas, R.C. (2016) Efficacy of occupational therapy task-oriented approach in upper extremity post-stroke rehabilitation. *Occupational Therapy International*, 23(4), 444-456.

Altschuler, E.L., Wisdom, S.B., Stone, L., Foster, C., Galasko, D., Llewellyn, D.M.E. and Ramachandran, V.S. (1999) Rehabilitation of hemiparesis after stroke with a mirror. *The Lancet*, 353(9169), 2035-2036.

American Occupational Therapy Association. (2014) Occupational therapy practice framework: Domain and process (3rd ed.). *American Journal of Occupational Therapy*, 68(Suppl. 1), S1-S48.

Anandabai, J. and Gupta, M. (2013) Effects of bimanual functional practice training versus unimanual functional practice training on functional performance of upper extremity in chronic stroke. *Journal of Physiotherapy and Sports Medicine*, 2(1), 15-30.

Angst, F., Goldhahn, J., Drerup, S., Aeschlimann, A., Schwyzer, H. K., and Simmen, B. R. (2008) Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. *Arthritis Care and Research*, 59(3), 391-398.

Arain, M., Campbell, M.J., Cooper, C.L. and Lancaster, G.A., 2010 What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology*, 10(1), 67.

Arya, K.N., Pandian, S., Verma, R. and Garg, R.K. (2011) Movement therapy induced neural reorganization and motor recovery in stroke: a review. *Journal of Bodywork and Movement Therapies*, 15(4), 528-537.

Arya, N.K., Verma, R., Garg, R.K., Sharma, V.P., Agarwal, M. and Aggarwal, G.G. (2012) Meaningful task-specific training (MTST) for stroke rehabilitation: a randomized controlled trial. *Topics in Stroke Rehabilitation*, 19(3), 193-211.

Ashford, S., Slade, M., Malaprade, F. and Turner-Stokes, L. (2008) Evaluation of functional outcome measures for the hemiparetic upper limb: a systematic review. *Journal of Rehabilitation Medicine*, 40(10), 787-795.

Åstrand, A., Saxin, C., Sjöholm, A., Skarin, M., Linden, T., Stoker, A., Roshandel, S., Dedering, Å., Halvorsen, M., Bernhardt, J. and Cumming, T. (2016) Poststroke physical activity levels no higher in rehabilitation than in the acute hospital. *Journal of Stroke and Cerebrovascular Diseases*, 25(4), 938-945.

Baker, K., Cano, S.J. and Playford, E.D. (2011) Outcome measurement in stroke. *Stroke,* 42(6), 1787-1794.

Barker, R.N. and Brauer, S.G. (2005) Upper limb recovery after stroke: the stroke survivors' perspective. *Disability and Rehabilitation*, 27(20), 1213-1223.

Bartels, M.N., Duffy, C.A. and Beland, H.E. (2016) Pathophysiology, medical management, and acute rehabilitation of stroke survivors. *In*: Gillen G, ed. *Stroke Rehabilitation: A Function-Based Approach*. 4th ed. Missouri: Elsevier, 2-45.

Beaton, D. E. (2000) Understanding the relevance of measured change through studies of responsiveness. *Spine*, 25(24), 3192-3199.

Béjot, Y., Bailly, H., Durier, J. and Giroud, M. (2016) Epidemiology of stroke in Europe and trends for the 21st century. *La Presse Médicale*, 45(12), e391-e398.

Beninato, M., Gill-Body, K. M., Salles, S., Stark, P. C., Black-Schaffer, R. M., and Stein, J. (2006) Determination of the minimal clinically important difference in the FIM instrument in patients with stroke. *Archives of Physical Medicine and Rehabilitation*, 87(1), 32-39.

Black, N. (2013) Patient reported outcome measures could help transform healthcare. *BMJ*, 346, f167.

Bland, J.M. and Altman, D. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307-310.

Bland, J.M. and Altman, D.G. (1999) Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160.

Bodenheimer, T., Wagner, E.H. and Grumbach, K. (2002) Improving primary care for patients with chronic illness. *Jama*, 288(14), 1775-1779.

Bodiam, C. (1999) The use of the Canadian Occupational Performance Measure for the assessment of outcome on a neurorehabilitation unit. *British Journal of Occupational Therapy*, 62(3), 123-126.

Bogard, K., Wolf, S., Zhang, Q., Thompson, P., Morris, D. and Nichols-Larsen, D. (2009) Can the wolf motor function test be streamlined? *Neurorehabilitation and Neural Repair*, 23(5), 422-428.

Bonifer, N. and Anderson, K.M. (2003) Application of constraint-induced movement therapy for an individual with severe chronic upper-extremity hemiplegia. *Physical Therapy*, 83(4), 384–398.

Bonifer, N.M., Anderson, K.M. and Arciniegas, D.B. (2005) Constraint-induced movement therapy after stroke: efficacy for patients with minimal upper-extremity motor ability. Archives of Physical Medicine and Rehabilitation, 86(9), 1867–1873.

Bottemiller, K. L., Bieber, P. L., Basford, J. R., and Harris, M. (2006) FIM scores, FIM efficiency, and discharge disposition following inpatient stroke rehabilitation. *Rehabilitation Nursing*, 31(1), 22-25.

Bourke, B. (2014) Positionality: Reflecting on the research process. *The Qualitative Report*, 19(33), 1-9.

Bowman, M.H., Taub, E., Uswatte, G., Delgado, A., Bryson, C., Morris, D.M., McKay, S. and Mark, V.W. (2006) A treatment for a chronic stroke patient with a plegic hand combining CI therapy with conventional rehabilitation procedures: case report. *NeuroRehabilitation*, 21(2), 167-176.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

British Heart Foundation. (2020) *Northern Ireland Factsheet January 2020*. Belfast: British Heart Foundation.

Buma, F., Kwakkel, G. and Ramsey, N. (2013) Understanding upper limb recovery after stroke. *Restorative Neurology and Neuroscience*, 31(6), 707-722.

Bushnell, C., Bettger, J.P., Cockroft, K.M., Cramer, S.C., Edelen, M.O., Hanley, D., Katzan, I.L., Mattke, S., Nilsen, D.M., Piquado, T. and Skidmore, E.R. (2015) Chronic stroke outcome measures for motor function intervention trials: expert panel recommendations. *Circulation: Cardiovascular Quality and Outcomes*, 8(6_suppl_3), S163-S169.

Cano, S. J., O'Connor, R. J., Thompson, A. J., and Hobart, J. C. (2006) Exploring disability rating scale responsiveness II: do more response options help? *Neurology*, 67(11), 2056-2059.

Cantero-Téllez, R., Naughton, N., Algar, L., and Valdes, K. (2019) Outcome measurement of hand function following mirror therapy for stroke rehabilitation: a systematic review. *Journal of Hand Therapy*, 32(2), 277-291.

Carod-Artal, F.J. and Egido, J.A. (2009) Quality of life after stroke: the importance of a good recovery. *Cerebrovascular Diseases*, 27(S1), 204-214.

Carswell, A., McColl, M.A., Baptiste, S., Law, M., Polatajko, H. and Pollock, N. (2004) The Canadian Occupational Performance Measure: a research and clinical literature review. *Canadian Journal of Occupational Therapy*, 71(4), 210-222.

Ch'Ng, A.M., French, D. and Mclean, N. (2008) Coping with the challenges of recovery from stroke: long term perspectives of stroke support group members. *Journal of Health Psychology*, 13(8), 1136-1146.

Chen, P., Lin, K.C., Liing, R.J., Wu, C.Y., Chen, C.L. and Chang, K.C. (2016) Validity, responsiveness, and minimal clinically important difference of EQ-5D-5L in stroke patients undergoing rehabilitation. *Quality of Life Research*, 25(6), 1585-1596.

Chen, Y.H., Rodger, S. and Polatajko, H. (2002) Experiences with the COPM and client-centred practice in adult neurorehabilitation in Taiwan. *Occupational Therapy International*, 9(3), 167-184.

Cicchetti, D. V. (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.

Clarke, D.J., Tyson, S., Rodgers, H., Drummond, A., Palmer, R., Prescott, M., Tyrrell, P., Burton, L., Grenfell, K., Brkic, L. and Forster, A. (2015) Why do patients with stroke not receive the recommended amount of active therapy (ReAcT)? Study protocol for a multisite case study investigation. *BMJ Open*, 5(8), e008443.

Cohen, D. J. and Crabtree, B. F. (2008) Evaluative criteria for qualitative research in health care: controversies and recommendations. *The Annals of Family Medicine*, 6(4), 331-339.

Cohen, M.E. and Marino, R.J. (2000) The tools of disability outcomes research functional status measures. *Archives of Physical Medicine and Rehabilitation*, 81(S2), S21-S29.

Coleman, K., Austin, B.T., Brach, C. and Wagner, E.H. (2009) Evidence on the chronic care model in the new millennium. *Health Affairs*, 28(1), 75-85.

College of Occupational Therapists. (2015) *Code of Ethics and Professional Conduct*. London: College of Occupational Therapists.

College of Occupational Therapists (2017) Professional Standards for Occupational Therapy Practice. London: College of Occupational Therapists.

Colquhoun, H., Letts, L., Law, M., MacDermid, J. and Edwards, M. (2010) Feasibility of the Canadian Occupational Performance Measure for routine use. *British Journal of Occupational Therapy*, 73(2), 48-54.

Constraint Induced Movement Therapy Research Group. (2000) *Manual: Graded Wolf Motor Function Test.* Birmingham, Ala: University of Alabama and Birmingham Veteran's Administration Centre.

Constraint Induced Movement Therapy Research Group. (2002) *Manual: Graded Wolf Motor Function Test*. Birmingham: University of Alabama and Birmingham Veteran's Administration Centre.

Coster, W.J. (2013) Making the best match: selecting outcome measures for clinical trials and outcome studies. *American Journal of Occupational Therapy*, 67(2), 162-170.

Coupar, F., Pollock, A., Rowe, P., Weir, C. and Langhorne, P. (2012) Predictors of upper limb recovery after stroke: a systematic review and meta-analysis. *Clinical Rehabilitation*, 26(4), 291-313.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, 337, a1655.

Cramer, S.C. (2008) Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. *Annals of Neurology*, 63(3), 272-287.

Cramer, S.C., Sur, M., Dobkin, B.H., O'brien, C., Sanger, T.D., Trojanowski, J.Q., Rumsey, J.M., Hicks, R., Cameron, J., Chen, D. and Chen, W.G. (2011) Harnessing neuroplasticity for clinical applications. *Brain*, 134(6), 1591-1609.

Cramer, S.C., Wolf, S.L., Adams Jr, H.P., Chen, D., Dromerick, A.W., Dunning, K., Ellerbe, C., Grande, A., Janis, S., Lansberg, M.G. and Lazar, R.M. (2017) Stroke recovery and rehabilitation research: issues, opportunities, and the National Institutes of Health StrokeNet. *Stroke*, 48(3), 813-819.

Crawford, M.J., Rutter, D., Manley, C., Weaver, T., Bhui, K., Fulop, N. and Tyrer, P. (2002) Systematic review of involving patients in the planning and development of health care. *BMJ*, 325(7375), 1263.

Critical Appraisal Skills Programme (2018) *CASP cohort study checklist [online]*. Available at: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Diagnostic-Checklist-2018.pdf [accessed 11 August 2018].

Crosby, R. D., Kolotkin, R. L., and Williams, G. R. (2003) Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, 56(5), 395-407.

Cup, E.H., Scholte op Reimer, W.J.M., Thijssen, M.C. and van Kuyk-Minis, M.A.H. (2003) Reliability and validity of the Canadian Occupational Performance Measure in stroke patients. *Clinical Rehabilitation*, 17(4), 402-409.

Deconinck, F.J.A., Smorenburg, A.R.P., Benham, A., Ledebt, A., Feltham, M.G. And Savelsbergh, G.J.P. (2015) Reflections on mirror therapy: a systematic review of the effect of mirror visual feedback on the brain. *Neurorehabilitation and Neural Repair,* 29(4), 349-361.

Davy, C., Bleasel, J., Liu, H., Tchan, M., Ponniah, S. and Brown, A. (2015) Effectiveness of chronic care models: opportunities for improving healthcare practice and health outcomes: a systematic review. *BMC Health Services Research*, 15(194).

Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., and Carr, A. J. (2010) The routine use of patient reported outcome measures in healthcare settings. *BMJ*, 340(c186).

De Savigny, D. and Adam, T. (2009) Systems thining for healh systems strengthening: an introduction. *In*: de Savigny, D. and Adam, T. eds. *Systems thinking for health systems strengthening*. Geneva: WHO. Available at: https://apps.who.int/iris/bitstream/handle/10665/44204/9789241563895_eng.pdf?sequence=1&isAllowed=y [Accessed 5 April 2020].

De Vet, H.C., Terwee, C.B., Knol, D.L. and Bouter, L.M. (2006) When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, 59(10), 1033-1039.

Demirtas-Tatlidede, A., Alonso-Alonso, M., Shetty, R.P., Ronen, I., Pascual-Leone, A. and Fregni, F. (2015) Long-term effects of contralesional rTMS in severe stroke: safety, cortical excitability, and relationship with transcallosal motor fibers. *NeuroRehabilitation*, 36(1), 51-59.

Department of Health. (2008) *Guidance on the routine collection of patient reported outcome measures (PROMs) for the NHS in England 2009/10*. London: The Stationary Office.

Department of Health. (2010) *Equity and excellence: liberating the NHS*. (Cm 7881). London: The Stationery Office.

Department of Health. (2019) *Reshaping stroke care saving lives, reducing disability. Consultation document*. Belfast: Department of Health.

Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B. and van Hout, B. (2018) Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics*, 27(1), 7-22.

Devlin, N. J., and Appleby, J. (2010) *Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision-making*. London: The King's Fund.

Deyo, R. A., and Centor, R. M. (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *Journal of Chronic Diseases*, 39(11), 897-906.

DeVellis, R. F. (2006) Classical test theory. *Medical care*, 44(11), S50-S59.

Doody, O., Slevin, E., and Taggart, L. (2013) Focus group interviews in nursing research: part 1. *British Journal of Nursing*, 22(1), 16-19.

Donnelly, C., O'Neill, C., Bauer, M. and Letts, L. (2017) Canadian Occupational Performance Measure (COPM) in primary care: A profile of practice. *American Journal of Occupational Therapy*, 71(6), 7106265010p1-7106265010p8.

Duff, S.V., He, J., Nelsen, M.A., Lane, C.J., Rowe, V.T., Wolf, S.L., Dromerick, A.W. and Winstein, C.J. (2015) Interrater reliability of the Wolf Motor Function Test-Functional Ability Scale: why it matters. *Neurorehabilitation and Neural Repair*, 29(5), 436-443.

Duncan Millar, J., van Wijck, F., Pollock, A. and Ali, M. (2019) Outcome measures in post-stroke arm rehabilitation trials: do existing measures capture outcomes that are important to stroke survivors, carers, and clinicians? *Clinical Rehabilitation*, 33(4), 737-749.

Dyer, M.T., Goldsmith, K.A., Sharples, L.S. and Buxton, M.J., 2010. A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health and Quality of Life Outcomes*, 8(1), p.13.

Edwards, D.F., Lang, C.E., Wagner, J.M., Birkenmeier, R. and Dromerick, A.W. (2012) An evaluation of the Wolf Motor Function Test in motor trials early after stroke. *Archives of Physical Medicine and Rehabilitation*, 93(4), 660-668.

Eldridge, S.M., Lancaster, G.A., Campbell, M.J., Thabane, L., Hopewell, S., Coleman, C.L. and Bond, C.M. (2016a) Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PloS one*, 11(3), e0150205.

Eldridge, S.M., Chan, C.L., Campbell, M.J., Bond, C.M., Hopewell, S., Thabane, L. and Lancaster, G.A. (2016b) CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ*, 355(i5239).

Enemark Larsen, A., Jessen Winge, C. and Christensen, J.R. (2019) Clinical utility of the Danish version of the Canadian Occupational Performance Measure. *Scandinavian Journal of Occupational Therapy*, In Press. Available at: https://www.tandfonline.com/doi/pdf/10.1080/11038128.2019.1634150?needAccess=true [Accessed 13 February 2020].

Etikan, I., Musa, S. A., and Alkassim, R. S. (2016) Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.

Essery, R., Geraghty, A.W., Kirby, S. and Yardley, L. (2017) Predictors of adherence to home-based physical therapies: a systematic review. *Disability and Rehabilitation*, 39(6), 519-534.

EuroQol Research Foundation. (2019) *EQ-5D-5L user guide, 2019*. Rotterdam: EuroQol Research Foundation. Available at: https://euroqol.org/publications/user-guides.

EuroQol Research Foundation. (2020) *EQ-5D-5L | Valuation: Standard value sets*. Rotterdam: EuroQol Research Foundation. Available at: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/ [Accessed 30 April 2020].

Evans, D. (2003) Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, 12(1), 77–84.

Eyssen, I.C., Steultjens, M.P., Oud, T.A., Bolt, E.M., Maasdam, A. and Dekker, J. (2011) Responsiveness of the Canadian Occupational Performance Measure. *Journal of Rehabilitation Research and Development*, 48(5), 517-528.

Ferlie, E.B. and Shortell, S.M. (2001) Improving the quality of health care in the United Kingdom and the United States: a framework for change. *The Milbank Quarterly*, 79(2), 281-315.

Fischer, H.C., Triandafilou, K.M., Thielbar, K.O., Ochoa, J.M., Lazzaro, E.D., Pacholski, K.A. and Kamper, D.G. (2015) Use of a portable assistive glove to facilitate rehabilitation in stroke survivors with severe hand impairment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(3), 344-351.

Fletcher, B., Gheorghe, A., Moore, D., Wilson, S. and Damery, S. (2012) Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. *BMJ Open*, 2(1), e000496.

Flinn, N.A., Smith, J.L., Tripp, C.J. and White, M.W. (2009) Effects of robotic-aided rehabilitation on recovery of upper extremity function in chronic stroke: a single case study. *Occupational Therapy International*, 16(3-4), 232-243.

Franceschini, M., La, F.P., Agosti, M. and Massucci, M. (2010) Is health-related-quality of life of stroke patients influenced by neurological impairments at one year after stroke? *European Journal of Physical and Rehabilitation Medicine*, 46(3), 389-399.

Fritz, S.L., Blanton, S., Uswatte, G., Taub, E. and Wolf, S.L. (2009) Minimal detectable change scores for the Wolf Motor Function Test. *Neurorehabilitation and Neural Repair*, 23(7), 662-667.

Galvin, R., Cusack, T., O'Grady, E., Murphy, T.B. and Stokes, E. (2011) Family-mediated exercise intervention (FAME) evaluation of a novel form of exercise delivery after stroke. *Stroke*, 42(3), 681-686.

GBD 2016 Stroke Collaborators. (2019) Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 18(5), 439-458.

GBD 2017 DALYs and HALE Collaborators. (2018) Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392(10159), 1859-1922.

Gillen, G. and Nilsen, D. M. (2016) Upper extremity function and management. *In*: Gillen, G. ed. *Stroke rehabilitation: a function-based approach*. 4th ed. Missouri: Elsevier, 424-485.

Gladstone, D.J., Danells, C.J. and Black, S.E. (2002) The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabilitation and Neural Repair*, 16(3), 232-240.

Glass, T.A. and Maddox, G.L. (1992) The quality and quantity of social support: stroke recovery as psycho-social transition. *Social Science and Medicine*, 34(11), 1249-1261.

Golicki, D., Niewada, M., Karlińska, A., Buczek, J., Kobayashi, A., Janssen, M.F. and Pickard, A.S. (2015a) Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ VAS in stroke patients. *Quality of Life Research*, 24(6), 1555-1563.

Golicki, D., Niewada, M., Buczek, J., Karlińska, A., Kobayashi, A., Janssen, M.F. and Pickard, A.S. (2015b) Validity of EQ-5D-5L in stroke. *Quality of Life Research*, 24(4), 845-850.

Guyatt, G., Walter, S., and Norman, G. (1987) Measuring change over time: assessing the usefulness of evaluative instruments. Journal of Clinical *Epidemiology*, 40(2), 171-178.

Hallgren, K. A. (2012) Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34.

Ham, C., Heenan, D., Longley, M. and Steel, D.R. (2013) *Integrated care in Northern Ireland, Scotland and Wales. Lessons for England*. London: The King's Fund.

Hanley, J. A., and McNeil, B. J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.

Harris, T. J., Carey, I. M., Victor, C. R., Adams, R., and Cook, D. G. (2008) Optimising recruitment into a study of physical activity in older people: a randomised controlled trial of different approaches. *Age and Ageing*, 37(6), 659-665.

Harris, J.E., Eng, J.J., Miller, W.C. and Dawson, A.S. (2010) The role of caregiver involvement in upper-limb treatment in individuals with subacute stroke. *Physical Therapy*, 90(9), 1302-1310.

Harrison, J.K., McArthur, K.S. and Quinn, T.J. (2013) Assessment scales in stroke: clinimetric and clinical considerations. *Clinical Interventions in Aging*, 8, 201-211.

Hatano S. (1976) Experience from a multicentre stroke register: a preliminary report. *Bulletin of the World Health Organisation*, 54(5), 541–553.

Heneghan, C., Goldacre, B. and Mahtani, K.R. (2017) Why clinical trial outcomes fail to translate into benefits for patients. *Trials*, 18(1), 122.

Hernaez, R. (2015) Reliability and agreement studies: a guide for clinical investigators. *Gut*, 64(7), 1018-1027.

Higgins, J.P., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L. and Sterne, J.A. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928.

Hobart, J.C., Lamping, D.L., Freeman, J.A., Langdon, D.W., McLellan, D.L., Greenwood, R.J. and Thompson, A.J. (2001) Evidence-based measurement: which disability scale for neurologic rehabilitation? *Neurology*, 57(4), 639-644.

Hobart, J. C., Cano, S. J., Zajicek, J. P., and Thompson, A. J. (2007) Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *The Lancet Neurology*, 6(12), 1094-1105.

Hodics, T.M., Nakatsuka, K., Upreti, B., Alex, A., Smith, P.S. and Pezzullo, J.C. (2012) Wolf Motor Function Test for characterizing moderate to severe hemiparesis in stroke patients. *Archives of Physical Medicine and Rehabilitation*, 93(11), 1963-1967.

Hole, E., Stubbs, B., Roskell, C. and Soundy, A. (2014) The patient's experience of the psychosocial process that influences identity following stroke rehabilitation: a metaethnography. *The Scientific World Journal*, 349151.

Hopman, W.M. and Verner, J. (2003) Quality of life during and after inpatient stroke rehabilitation. *Stroke*, 34(3), 801-805.

Hopman, P., De Bruin, S.R., Forjaz, M.J., Rodriguez-Blazquez, C., Tonnara, G., Lemmens, L.C., Onder, G., Baan, C.A. and Rijken, M. (2016) Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty: a systematic literature review. *Health Policy*, 120(7), 818-832.

Horsburgh, D. (2003) Evaluation of qualitative research. *Journal of Clinical Nursing*, 12(2), 307-312.

Houwink, A., Nijland, R.H., Geurts, A.C. and Kwakkel, G. (2013) Functional recovery of the paretic upper limb after stroke: who regains hand capacity? *Archives of Physical Medicine and Rehabilitation,* 94(5), 839-844.

Horne, M., Thomas, N., McCabe, C., Selles, R., Vail, A., Tyrrell, P. and Tyson, S. (2015a) Patient-directed therapy during in-patient stroke rehabilitation: stroke survivors' views of feasibility and acceptability. *Disability and Rehabilitation*, 37(25), 2344-2349.

Horne, M., Thomas, N., Vail, A., Selles, R., McCabe, C. and Tyson, S. (2015b) Staff's views on delivering patient-led therapy during inpatient stroke rehabilitation: a focus group study with lessons for trial fidelity. *Trials*, 16(1), 137.

Hsieh, Y.W., Wu, C.Y., Lin, K.C., Chang, Y.F., Chen, C.L. and Liu, J.S. (2009) Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. *Stroke*, 40(4), 1386-1391.

Hsieh, Y.W., Chang, K.C., Hung, J.W., Wu, C.Y., Fu, M.H. and Chen, C.C. (2018) Effects of home-based versus clinic-based rehabilitation combining mirror therapy and task-specific training for patients with stroke: a randomized crossover trial. *Archives of Physical Medicine and Rehabilitation*, 99(12), 2399-2407.

Hubbard, G., O'Carroll, R., Munro, J., Mutrie, N., Haw, S., Mason, H. and Treweek, S. (2016) The feasibility and acceptability of trial procedures for a pragmatic randomised controlled trial of a structured physical activity intervention for people diagnosed with

colorectal cancer: findings from a pilot trial of cardiac rehabilitation versus usual care (no rehabilitation) with an embedded qualitative study. *Pilot and Feasibility Studies*, 2(1), 51.

Husted, J. A., Cook, R. J., Farewell, V. T., and Gladman, D. D. (2000) Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology*, 53(5), 459-468.

Inouye, M., Hashimoto, H., Mio, T. and Sumino, K. (2001) Influence of admission functional status on functional change after stroke rehabilitation. *American Journal of Physical Medicine and Rehabilitation*, 80(2), 121-125.

Institute of Medicine. (2001) *Crossing the quality chasm: A new health system for the 21st century*. Washington: National Academy Press.

Intercollegiate Stroke Working Party. (2016) *National Clinical Guideline for Stroke*. 5th ed. London: Royal College of Physicians.

Ivey, J. and Mew, M. (2010) Theoretical basis. *In*: Edmans, J. ed. *Occupational therapy and stroke*. 2nd ed. Wiley-Blackwell: West Sussex, 24-48.

Iwamuro, B.T., Fischer, H.C. and Kamper, D.G. (2011) A pilot study to assess use of passive extension bias to facilitate finger movement for repetitive task practice after stroke. *Topics in Stroke Rehabilitation*, 18(4), 308–315.

Janssen, M.F., Pickard, A.S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P. and Busschbach, J. (2013) Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*, 22(7), 1717-1727.

Jerosch-Herold, C. (2005) An evidence-based approach to choosing outcome measures: a checklist for the critical appraisal of validity, reliability and responsiveness studies. *British Journal of Occupational Therapy*, 68(8), 347–353.

Jin, Y. P., Di Legge, S., Ostbye, T., Feightner, J. W., and Hachinski, V. (2006) The reciprocal risks of stroke and cognitive impairment in an elderly population. *Alzheimer's and Dementia*, 2(3), 171-178.

Jokinen, H., Melkas, S., Ylikoski, R., Pohjasvaara, T., Kaste, M., Erkinjuntti, T. and Hietanen, M. (2015) Post-stroke cognitive impairment is common even after successful clinical recovery. *European Journal of Neurology*, 22(9), 1288-1294.

Jones, G. (2018) Raising the profile of pilot and feasibility studies in relation to the development, evaluation and implementation of patient-reported outcome measures. *Pilot and Feasibility Studies*, 4(3).

Jurkiewicz, M.T., Marzolini, S. and Oh, P. (2011) Adherence to a home-based exercise program for individuals after stroke. *Topics in Stroke Rehabilitation*, 18(3), 277-284.

Kalra, L., Evans, A., Perez, I., Melbourn, A., Patel, A., Knapp, M. and Donaldson, N., (2004) Training carers of stroke patients: randomised controlled trial. *BMJ*, 328(7448), 1099.

Kamper, S.J., Maher, C.G. and Mackay, G. (2009) Global rating of change scales: a review of strengths and weaknesses and considerations for design. *Journal of Manual and Manipulative Therapy*, 17(3), 163-170.

Kessler, D., Egan, M., Dubouloz, C.J., McEwen, S. and Graham, F.P. (2017) Occupational performance coaching for stroke survivors: a pilot randomized controlled trial. *American Journal of Occupational Therapy*, 71(3), 7103190020p1-7103190020p7.

Koo, T.K. and Li, M.Y. (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.

Kottner, J. and Dassen, T. (2008) An interrater reliability study of the Braden scale in two nursing homes. *International Journal of Nursing Studies*, 45(10), 1501-1511.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M. and Streiner, D.L. (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1), 96-106.

Krishnamurthi, R.V., Moran, A.E., Feigin, V.L., Barker-Collo, S., Norrving, B., Mensah, G.A., Taylor, S., Naghavi, M., Forouzanfar, M.H., Nguyen, G., Johnson, C.O., Vos, T, Murray, C.J.L. and Roth, G.A. for the GBD 2013 Stroke Panel Experts Group. (2015) Stroke prevalence, mortality and disability-adjusted life years in adults aged 20-64 years in 1990-2013: data from the global burden of disease 2013 study. *Neuroepidemiology*, 45(3), 190-202.

Krakauer, J.W., Carmichael, S.T., Corbett, D. and Wittenberg, G.F. (2012) Getting neurorehabilitation right: what can be learned from animal models? *Neurorehabilitation and Neural Repair*, 26(8), 923-931.

Kwakkel, G., Kollen, B.J., van der Grond, J. and Prevo, A.J. (2003) Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke*, 34(9), 2181-2186.

Kwakkel, G., and Kollen, B. J. (2013) Predicting activities after stroke: what is clinically relevant? *International Journal of Stroke*, 8(1), 25-32.

Kwakkel, G., Lannin, N.A., Borschmann, K., English, C., Ali, M., Churilov, L., Saposnik, G., Winstein, C., Van Wegen, E.E., Wolf, S.L. and Krakauer, J.W. (2017) Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. *Neurorehabilitation and Neural Repair*, 31(9), 784-792.

Lancaster, G.A., Dodd, S. and Williamson, P.R. (2004) Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice*, 10(2), 307-312.

Langhorne, P., Coupar, F. and Pollock, A. (2009) Motor recovery after stroke: a systematic review. *The Lancet Neurology*, 8(8), 741-754.

Langhorne, P., Bernhardt, J. and Kwakkel, G. (2011) Stroke rehabilitation. *The Lancet*, 377(9778), 1693-1702.

Langhorne, P., Ramachandra, S., and Stroke Unit Trialists' Collaboration (2020). Organised inpatient (stroke unit) care for stroke: network meta-analysis. Cochrane Database of Systematic Reviews, Issue 4. Art No.: CD000197. DOI:10.1002/14651858.CD000197.pub4

Lang, C. E., Edwards, D. F., Birkenmeier, R. L., and Dromerick, A. W. (2008) Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Archives of Physical Medicine and Rehabilitation*, 89(9), 1693-1700.

Lang, C.E., MacDonald, J.R., Reisman, D.S., Boyd, L., Kimberley, T.J., Schindler-Ivens, S.M., Hornby, T.G., Ross, S.A. and Scheets, P.L. (2009) Observation of amounts of movement practice provided during stroke rehabilitation. *Archives of Physical Medicine and Rehabilitation*, 90(10), 1692-1698.

Lang, C.E., Bland, M.D., Bailey, R.R., Schaefer, S.Y. and Birkenmeier, R.L. (2013) Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. *Journal of Hand* Therapy, 26(2), 104-115.

Law, M., Baptiste, S., Carswell, A., McColl, M.A., Polatajko, H. and Pollock, N. (2014) *Canadian Occupational Performance Measure*. 5th ed. Ottawa: CAOT Publications ACE.

Lawrence, E.S., Coshall, C., Dundas, R., Stewart, J., Rudd, A.G., Howard, R. and Wolfe, C.D. (2001) Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke,* 32(6), 1279-1284.

Legg, L.A., Lewis, S.R., Schofield-Robinson, O.J., Drummond, A. and Langhorne, P. (2017) Occupational therapy for adults with problems in activities of daily living after stroke. *Cochrane Database of Systematic Reviews*, Issue 7. Art. No.: CD003585. DOI: 10.1002/14651858.CD003585.pub3.

Lewin, S., Glenton, C. and Oxman, A.D. (2009) Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *BMJ*, 339, b3496.

Liang, M. H., Fossel, A. H., and Larson, M. G. (1990) Comparisons of five health status instruments for orthopedic evaluation. *Medical Care*, 28(7), 632-642.

Lin, J.H., Hsu, M.J., Sheu, C.F., Wu, T.S., Lin, R.T., Chen, C.H. and Hsieh, C.L. (2009a) Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Physical Therapy*, 89(8), 840-850.

Lin, K.C., Hsieh, Y.W., Wu, C.Y., Chen, C.L., Jang, Y. and Liu, J.S. (2009b) Minimal detectable change and clinically important difference of the Wolf Motor Function Test in stroke patients. *Neurorehabilitation and Neural Repair*, 23(5), 429-434.

Lu, W.S., Huang, S.L., Yang, J.F., Chen, M.H., Hsieh, C.L. and Chou, C.Y. (2016) Convergent validity and responsiveness of the EQ-5D utility weights for stroke survivors. *Journal of Rehabilitation Medicine*, 48(4), 346-351.

Luker, J., Lynch, E., Bernhardsson, S., Bennett, L. and Bernhardt, J. (2015) Stroke survivors' experiences of physical rehabilitation: a systematic review of qualitative studies. *Archives of Physical Medicine and Rehabilitation*, 96(9), 1698-1708.

Maclean, N., Pound, P., Wolfe, C. and Rudd, A. (2000) Qualitative analysis of stroke patients' motivation for rehabilitation. BMJ, 321(7268), 1051-1054.

Madden, S., Hopman, W.M., Bagg, S., Verner, J. and O'Callaghan, C.J. (2006) Functional status and health-related quality of life during inpatient stroke rehabilitation. *American Journal of Physical Medicine and Rehabilitation*, 85(10), 831-838.

Maeshima, S., Ueyoshi, A., Osawa, A., Ishida, K., Kunimoto, K., Shimamoto, Y., Matsumoto, T. and Yoshida, M. (2003) Mobility and muscle strength contralateral to

hemiplegia from stroke: benefit from self-training with family support. *American Journal of Physical Medicine and Rehabilitation*, 82(6), 456-462.

McCabe, C. (2011) Mirror visual feedback therapy. A practical approach. *Journal of Hand Therapy*, 24(2), 170-179.

McHorney, C. A., and Tarlov, A. R. (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research*, 4(4), 293–307.

McLellan, E., MacQueen, K.M. and Neidig, J.L. (2003) Beyond the qualitative interview: Data preparation and transcription. *Field Methods*, 15(1), 63-84.

Meadmore, K.L., Hallewell, E., Freeman, C. and Hughes, A.M. (2019) Factors affecting rehabilitation and use of upper limb after stroke: views from healthcare professionals and stroke survivors. *Topics in Stroke Rehabilitation*, 26(2), 94-100.

Middel, B. and Van Sonderen, E. (2002) Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*, 2(17), e15.

Miller, K.K., Porter, R.E., DeBaun-Sprague, E., Van Puymbroeck, M. and Schmid, A.A. (2017) Exercise after stroke: patient adherence and beliefs after discharge from rehabilitation. *Topics in Stroke Rehabilitation*, 24(2), 142-148.

Milne, J. and Oberle, K. (2005) Enhancing rigor in qualitative description. *Journal of Wound Ostomy and Continence Nursing*, 32(6), 413-420.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M. and de Vet, Henrica C. W. (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research,* 19(4), 539-549.

Mokkink, L.B., De Vet, H.C., Prinsen, C.A., Patrick, D.L., Alonso, J., Bouter, L.M. and Terwee, C.B. (2018) COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, 27(5), 1171-1179.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D. and Baird, J. (2015) Process evaluation of complex interventions: Medical Research Council guidance. BMJ, 350(h1258).

Moreland, J.D., DePaul, V.G., DeHueck, A.L., Pagliuso, S.A., Yip, D.W., Pollock, B.J. and Wilkins, S. (2009) Needs assessment of individuals with stroke after discharge from hospital stratified by acute Functional Independence Measure score. *Disability and Rehabilitation*, 31(26), 2185-2195.

Morkisch, N., Thieme, H., and Dohle, C. (2019) How to perform mirror therapy after stroke? Evidence from a meta-analysis. *Restorative Neurology and Neuroscience*, 37(5), 421-435.

Morris, D.M., Uswatte, G., Crago, J.E., Cook III, E.W. and Taub, E. (2001) The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Archives of Physical Medicine and Rehabilitation*, 82(6), 750-755.

Morris, D.M., Taub, E., Macrina, D.M., Cook, E.W. and Geiger, B.F. (2009) A method for standardising procedures in rehabilitation: Use in the extremity constraint induced therapy evaluation multi- site randomised controlled trial. *Archives of Physical Medicine and Rehabilitation*, 90(4), 663–668.

Morris, J.H., van Wijck, F., Joice, S. and Donaghy, M. (2013) Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. *Disability and Rehabilitation,* 35(4), 1-9.

Murphy, M.A., Resteghini, C., Feys, P. and Lamers, I. (2015) An overview of systematic reviews on upper extremity outcome measures after stroke. *BMC Neurology*, 15(1), 29.

Murphy, T.H. and Corbett, D. (2009) Plasticity during stroke recovery: from synapse to behaviour. *Nature Reviews Neuroscience*, 10(12), 861.

Nakayama, H., Jørgensen, H.S., Raaschou, H.O. and Olsen, T.S. (1994) Recovery of upper extremity function in stroke patients: the Copenhagen stroke study. *Archives of Physical Medicine and Rehabilitation*, *75*(4), 394-398.

Nasr, N., Leon, B., Mountain, G., Nijenhuis, S.M., Prange, G., Sale, P. and Amirabdollahian, F. (2016) The experience of living with stroke and using technology: opportunities to engage and co-design with end users. *Disability and Rehabilitation: Assistive Technology*, 11(8), 653-660.

National Board for Certified Occupational Therapists. (2008) *Executive summary for the practice analysis study.*
Available at: https://www.nbcot.org/-/media/NBCOT/PDFs/2017-Practice-Analysis-Executive-OTR.ashx?la=en. National Board for Certified Occupational Therapists: Gaithersburg. Accessed 7 November 2019.

National Institute for Health and Care Excellence. (2013) *Stroke rehabilitation: long-term rehabilitation in adults*. (Clinical guideline [CG162]). London: NICE.

National Institute for Health and Care Excellence. (2014) *Guide to the methods of technology appraisal 2013.* (Process and methods [PMG9]). London: NICE.

National Institute for Health and Care Excellence. (2019) Position statement on use of the EQ-5D-5L value set for England (updated October 2019). London: NICE. Available at: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l [Accessed 30 April 2020].

Needham, C. and Carr, S. (2009) *SCIE Research Briefing 31: Co-production: An emerging evidence base for adult social care transformation*. London: Social Care Institute for Excellence.

Neergaard, M. A., Olesen, F., Andersen, R. S., and Sondergaard, J. (2009) Qualitative description–the poor cousin of health research? *BMC Medical Research Methodology*, 9(1), 52.

Newington, L. and Metcalfe, A. (2014) Researchers' and clinicians' perceptions of recruiting participants to clinical research: a thematic meta-synthesis. *Journal of Clinical Medicine Research*, 6(3), 162-172.

Ng, Y.S., Astrid, S., De Silva, D.A., Tan, M.L.D., Tan, Y.L. and Chew, E. (2013) Functional outcomes after inpatient rehabilitation in a prospective stroke cohort. *Proceedings of Singapore Healthcare*, 22(3), 175-182.

Nichols-Larsen, D.S., Clark, P.C., Zeringue, A., Greenspan, A. and Blanton, S. (2005) Factors influencing stroke survivors' quality of life during subacute recovery. *Stroke*, 36(7), 1480-1484.

Nilsen, D.M., Gillen, G. and Gordon, A.M. (2010) Use of mental practice to improve upper-limb recovery after stroke: a systematic review. *American Journal of Occupational Therapy*, 64(5), 695-708.

Norman, G. R., Stratford, P., and Regehr, G. (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology*, 50(8), 869-879.

Norman, G.R., Wyrwich, K.W. and Patrick, D.L. (2007) The mathematical relationship among different forms of responsiveness coefficients. *Quality of Life Research*, 16(5), 815-822.

Northern Ireland Statistics and Research Agency. (2019) *Disease prevalence (quality outcomes framework)* (administrative geographies). Belfast: Department of Health. Available at: https://www.ninis2.nisra.gov.uk/public/PivotGrid.aspx?ds=9882&lh=73&yn=2007-2019&sk=134&sn=Health%20and%20Social%20Care&yearfilter= [Accessed 7 February 2020]

Nystuen, P., and Hagen, K. B. (2004) Telephone reminders are effective in recruiting nonresponding patients to randomized controlled trials. *Journal of Clinical Epidemiology*, 57(8), 773-776.

O'Cathain, A., Hoddinott, P., Lewin, S., Thomas, K.J., Young, B., Adamson, J., Jansen, Y.J., Mills, N., Moore, G. and Donovan, J.L. (2015) Maximising the impact of qualitative research in feasibility studies for randomised controlled trials: guidance for researchers. *Pilot and Feasibility Studies*, *1*(1), 32.

O'Donoghue, P. (2012) *Statistics for sport and exercise studies*. Abingdon: Routledge.

Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L. and Zoran, A.G. (2009) A qualitative framework for collecting and analyzing data in focus group research. *International Journal of Qualitative Methods*, *8*(3), 1-21.

Ostwald, S. K., Godwin, K. M., and Cron, S. G. (2009) Predictors of life satisfaction in stroke survivors and spousal caregivers after inpatient rehabilitation. *Rehabilitation Nursing*, *34*(4), 160-174.

Ottenbacher, K. J., Hsu, Y., Granger, C. V., and Fiedler, R. C. (1996) The reliability of the functional independence measure: a quantitative review. *Archives of Physical Medicine and Rehabilitation*, 77(12), 1226-1232.

Page, S. J., Fulk, G. D., and Boyne, P. (2012) Clinically important differences for the upper-extremity Fugl-Meyer Scale in people with minimal to moderate impairment due to chronic stroke. *Physical Therapy*, 92(6), 791-798.

Paget, S. P., Caldwell, P. H., Murphy, J., Lilischkis, K. J., and Morrow, A. M. (2017) Moving beyond 'not enough time': factors influencing paediatric clinicians' participation in research. *Internal Medicine Journal*, 47(3), 299-306.

Passalent, L.A., Tyas, J.E., Jaglal, S.B. and Cott, C.A. (2011) The FIM™ as a measure of change in function after discharge from inpatient rehabilitation: a Canadian perspective. *Disability and Rehabilitation*, 33(7), 579-588.

Pallesen, H. (2014) Body, coping and self-identity. A qualitative 5-year follow-up study of stroke. *Disability and Rehabilitation*, *36*(3), 232-241.

Pereira, N.D., Vieira, L., Pompeu, F.P., Menezes, I.D.S., Anjos, S.M.D. and Ovando, A.C. (2015) Translation, cultural adaptation and reliability of the Brazilian version of

the Graded Wolf Motor Function Test in adults with severe hemiparesis. *Fisioterapia em Movimento*, 28(4), 667-676.

Pickard, A. S., Johnson, J. A., and Feeny, D. H. (2005) Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research*, 14(1), 207-219.

Pike, S., Lannin, N.A., Wales, K. and Cusick, A. (2018) A systematic review of the psychometric properties of the Action Research Arm Test in neurorehabilitation. *Australian Occupational Therapy Journal*, 65(5), 449-471.

Pindus, D.M., Mullis, R., Lim, L., Wellwood, I., Rundell, A.V., Aziz, N.A.A. and Mant, J. (2018) Stroke survivors' and informal caregivers' experiences of primary care and community healthcare services– a systematic review and meta-ethnography. *PLoS One*, 13(2), e0192533.

Plummer-D'Amato, P. (2008) Focus group methodology Part 1: considerations for design. *International Journal of Therapy and Rehabilitation*, 15(2), 69-73.

Pollock, A., St George, B., Fenton, M. and Firkins, L. (2012) Top ten research priorities relating to life after stroke. *The Lancet Neurology*, 11(3), 209.

Pollock, A., Farmer Sybil E, Brady Marian C, Langhorne, P., Mead Gillian E, Mehrholz, J. and van Wijck, F. (2014) Interventions for improving upper limb function after stroke. *Cochrane Database of Systematic Reviews,* Issue 11. Art. No.: CD010820. doi: 10.1002/14651858.CD010820.pub2.

Price-Haywood, E.G., Harden-Barrios, J., Carr, C., Reddy, L., Bazzano, L.A. and van Driel, M.L. (2019) Patient-reported outcomes in stroke clinical trials 2002–2016: a systematic review. *Quality of Life Research*, 28(5), 1119-1128.

Proot, I.M., Crebolder, H.F., Abu-Saad, H.H., Macor, T.H. and Meulen, R.H.T. (2000) Facilitating and constraining factors on autonomy: The views of stroke patients on admission into nursing homes. *Clinical Nursing Research*, 9(4), 460-478.

Proot, I.M., Ter Meulen, R.H., Abu-Saad, H.H. and Crebolder, H.F. (2007) Supporting stroke patients' autonomy during rehabilitation. *Nursing Ethics*, 14(2), 229-241.

van der Putten, J.J.M.F., Hobart, J.C., Freeman, J.A. and Thompson, A.J. (1999) Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery and Psychiatry*, 66(4), 480-484.

Ramachandran, V.S. and Rogers-Ramachandran, D. (1996) Synaesthesia in phantom limbs induced with mirrors. *Proceedings: Biological Sciences,* 263(1369), 377-386.

Ramachandran, V.S. and Rogers-Ramachandran, D. (2000) Phantom limbs and neural plasticity. *Archives of Neurology*, 57(3), 317-320.

Ramachandran, V.S. and Altschuler, E.L. (2009) The use of visual feedback, in particular mirror visual feedback, in restoring brain function. *Brain,* 132(7), 1693-1710.

Reeve, B.B., Wyrwich, K.W., Wu, A.W., Velikova, G., Terwee, C.B., Snyder, C.F., Schwartz, C., Revicki, D.A., Moinpour, C.M., McLeod, L.D. and Lyons, J.C. (2013) ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research*, 22(8), 1889-1905.

Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., and Aaronson, N. K. (2006) Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*, 4(1), 70.

Revicki, D., Hays, R. D., Cella, D., and Sloan, J. (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102-109.

Reynolds, R., Dennis, S., Hasan, I., Slewa, J., Chen, W., Tian, D., Bobba, S. and Zwar, N. (2018) A systematic review of chronic disease management interventions in primary care. *BMC Family Practice*, 19**(**11).

Roe, D., Brown, T. and Thyer, L. (2019) Validity, responsiveness, and perceptions of clinical utility of the Canadian Occupational Performance Measure when used in a sub-acute setting. *Disability and Rehabilitation*. DOI: 10.1080/09638288.2019.1573934

Rosewilliam, S., Roskell, C.A. and Pandyan, A.D. (2011) A systematic review and synthesis of the quantitative and qualitative evidence behind patient-centred goal setting in stroke rehabilitation. *Clinical Rehabilitation*, 25(6), 501-514.

Rowland, T.J. and Gustafsson, L. (2008) Assessments of upper limb ability following stroke: a review. *British Journal of Occupational Therapy*, 71(10), 427-437.

Royal College of Physicians, Clinical Effectiveness and Evaluation Unit on behalf of the Intercollegiate Stroke Working Party. (2016) *Sentinel stroke national audit programme (SSNAP): clinical audit April-July 2016 public report. National results November 2016*. London: Royal College of Physicians.

Safaz, I., Ylmaz, B., Yasar, E. and Alaca, R. (2009) Brunnstrom recovery stage and motricity index for the evaluation of upper extremity in stroke: analysis for correlation and responsiveness. *International Journal of Rehabilitation Research*, 32(3), 228-231.

Saka, Ö., McGuire, A. and Wolfe, C. (2009) Cost of stroke in the United Kingdom. *Age and Ageing,* 38(1), 27-32.

Salinas, J., Sprinkhuizen, S.M., Ackerson, T., Bernhardt, J., Davie, C., George, M.G., Gething, S., Kelly, A.G., Lindsay, P., Liu, L. and Martins, S.C. (2016) An international standard set of patient-centered outcome measures after stroke. *Stroke*, 47(1), 180-186.

Salter, K., Hellings, C., Foley, N. and Teasell, R. (2008) The experience of living with stroke: a qualitative meta-synthesis. *Journal of Rehabilitation Medicine*, 40(8), 595-602.

Sandelowski, M. (2010) What's in a name? Qualitative description revisited. *Research in Nursing and Health*, *33*(1), 77-84.

Santisteban, L., Térémetz, M., Bleton, J., Baron, J., Maier, M.A. and Lindberg, P.G. (2016) Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. *PloS One,* 11(5), e0154792.

Satink, T., Cup, E.H., de Swart, B.J. and Nijhuis-van der Sanden, M.W. (2015) How is self-management perceived by community living people after a stroke? A focus group study. *Disability and Rehabilitation*, 37(3), 223-230.

Schepers, V. P. M., Ketelaar, M., Visser-Meily, J. M. A., Dekker, J., and Lindeman, E. (2006) Responsiveness of functional health status measures frequently used in stroke research. *Disability and Rehabilitation*, 28(17), 1035-1040.

Schiavi, M., Costi, S., Pellegrini, M., Formisano, D., Borghi, S. and Fugazzaro, S. (2018) Occupational therapy for complex inpatients with stroke: identification of occupational needs in post-acute rehabilitation setting. *Disability and Rehabilitation*, 40(9), 1026-1032.

Schulz, K.F., Altman, D.G. and Moher, D. (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials*, 11(1), 32.

Scorrano, M., Ntsiea, V. and Maleka, D. (2018) Enablers and barriers of adherence to home exercise programmes after stroke: caregiver perceptions. *International Journal of Therapy and Rehabilitation*, 25(7), 353-364.

See, J., Dodakian, L., Chou, C., Chan, V., McKenzie, A., Reinkensmeyer, D.J. and Cramer, S.C. (2013) A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. *Neurorehabilitation and Neural Repair*, 27(8), 732-741.

Sekhon, M., Cartwright, M. and Francis, J.J. (2017) Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Services Research*, 17(1), 88.

Sendall, M., McCosker, L., Crossley, K. and Bonner, A. (2017) A structured review of chronic care model components supporting transition between healthcare service delivery types for older people with multiple chronic diseases. *Health Information Management Journal*, 46(2), 58-68.

Shrout, P. E., and Fleiss, J. L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.

Sim, J. (1998) Collecting and analysing qualitative data: issues raised by the focus group. *Journal of Advanced Nursing*, 28(2), 345-352.

Singer, S.J., Burgers, J., Friedberg, M., Rosenthal, M.B., Leape, L. and Schneider, E. (2011) Defining and measuring integrated patient care: promoting the next frontier in health care delivery. *Medical Care Research and Review*, 68(1), 112-127.

Singer, S.J., Kerrissey, M., Friedberg, M. and Phillips, R. (2020) A comprehensive theory of integration. *Medical Care Research and Review*, 77(2), 196-207.

Sivan, M., O'Connor, R.J., Makower, S., Levesley, M. and Bhakta, B. (2011) Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. *Journal of Rehabilitation Medicine*, 43(3),181-189.

Smart, A. (2006) A multi-dimensional model of clinical utility. *International Journal for Quality in Health Care*, 18(5), 377-382.

Smith, C.T., Hickey, H., Clarke, M., Blazeby, J. and Williamson, P. (2014) The trials methodological research agenda: results from a priority setting exercise. *Trials,* 15(1), 32.

Spencer, L., Ritchie, J., Ormston, R., O'connor, W. and Barnard, M. (2014) Analysis: principles and processes. *In*: Ritchie, J., Lewis, J., Mcnaughton Nicholls, C. and Ormston, R. eds. *Qualitative Research Practice*. 2nd eds. London: Sage Publications Ltd. 269-294.

Stevens, E., Emmett, E., Wang, Y., McKevitt, C. and Wolfe, C. (2017) *The burden of stroke in Europe*. London: Stroke Alliance for Europe. Available at: www.strokeeurope.eu/downloads/TheBurdenOfStrokeInEuropeReport.pdf [Accessed 31 October 2019]

Stineman, M. G., Shea, J. A., Jette, A., Tassoni, C. J., Ottenbacher, K. J., Fiedler, R., and Granger, C. V. (1996) The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of Physical Medicine and Rehabilitation*, 77(11), 1101-1108.

Stratford, P. W., and Goldsmith, C. H. (1997) Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Physical Therapy*, 77(7), 745–750.

Streiner, D.L. (2003) Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99-103.

Streiner, D.L. (2010) Measure for measure: New developments in measurement and item response theory. *The Canadian Journal of Psychiatry*, 55(3), 180-186.

Streiner, D.L. and Kottner, J. (2014) Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, 70(9), 1970-1979.

Streiner, D.L., Norman, G.R. and Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. 5th ed. London: Oxford University Press.

Stroke Association. (2018) *State of the nation stroke statistics, February 2018*. London: Stroke Association. Available at: https://www.stroke.org.uk/sites/default/files/state_of_the_nation_2018.pdf [Accessed 31 October 2019]

Stroke Association. (2019) *Struggling to recover. Life after stroke in Northern Ireland*. London: Stroke Association.

Suenkeler, I.H., Nowak, M., Misselwitz, B., Kugler, C., Schreiber, W., Oertel, W.H. and Back, T. (2002) Timecourse of health-related quality of life as determined 3, 6 and 12 months after stroke. *Journal of Neurology*, 249(9), 1160-1167.

Sugavanam, T., Mead, G., Bulley, C., Donaghy, M. and van Wijck, F. (2013) The effects and experiences of goal setting in stroke rehabilitation–a systematic review. *Disability and Rehabilitation*, 35(3), 177-190.

Sullivan, J.E., Crowner, B.E., Kluding, P.M., Nichols, D., Rose, D.K., Yoshida, R. and Pinto Zipp, G. (2013) Outcome measures for individuals with stroke: process and recommendations from the American Physical Therapy Association neurology section task force. *Physical Therapy*, 93(10), 1383-1396.

Tatemichi, T. K., Desmond, D. W., Stern, Y., Paik, M., Sano, M., and Bagiella, E. (1994) Cognitive impairment after stroke: frequency, patterns, and relationship to functional abilities. *Journal of Neurology, Neurosurgery and Psychiatry*, 57(2), 202-207.

Taub E, Morris DM, Crago J, King, D.K., Bowman, M., Bryson, C., Bishop, S., Pearson, S. and Shaw, S.E. (2011) *Wolf Motor Function Test (WMFT) Manual*. Birmingham: UAB CI Therapy Research Group.

Taub, E., Uswatte, G., Mark, V.W., Morris, D.M., Barman, J., Bowman, M.H., Bryson, C., Delgado, A. and Bishop-McKay, S. (2013) Method for enhancing real-world use

of a more affected arm in chronic stroke: transfer package of constraint-induced movement therapy. *Stroke*, 44(5), 1383-1388.

Teasell, R., Mehta, S., Pereira, S., McIntyre, A., Janzen, S., Allen, L., Lobo, L. and Viana, R. (2012) Time to rethink long-term rehabilitation management of stroke patients. *Topics in Stroke Rehabilitation*, 19(6), 457-462.

Terwee, C.B., Dekker, F.W., Wiersinga, W.M., Prummel, M.F. and Bossuyt, P.M.M. (2003) On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research*, 12(4), 349-362.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M. and de Vet, H.C. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42.

The Regulation and Quality Improvement Authority. (2014) *The Regulation and Quality Improvement Authority review of stroke services in Northern Ireland December 2014*. Belfast: The Regulation and Quality Improvement Authority. Available at: https://www.rqia.org.uk/RQIA/files/b8/b8f067de-3bf7-40c6-9297-b21a41a31811.pdf [Accessed 15 April 2020].

Thieme, H., Mehrholz, J., Pohl, M., Behrens, J. and Dohle, C. (2012) Mirror therapy for improving motor function after stroke. *Cochrane Database of Systematic Reviews*, Issue 3. Art No.: CD008449. doi: 10.1002/14651858.CD008449.pub2.

Thieme, H., Morkisch N., Mehrholz, J., Pohl, M., Behrens, J., Borgetto, B., Dohle, C. (2018) Mirror therapy for improving motor function after stroke. Cochrane Database of Systematic Reviews, Issue 7. Art No.: CD008449. doi:10.1002/14651858.CD008449.pub3.

Thompson-Butel, A.G., Lin, G., Shiner, C.T. and McNulty, P.A. (2015) Comparison of three tools to measure improvements in upper-limb function with poststroke therapy. *Neurorehabilitation and Neural Repair*, 29(4), 341-348.

Thyer, L., Brown, T. and Roe, D. (2018) The validity of the Canadian Occupational Performance Measure (COPM) when used in a sub-acute rehabilitation setting with older adults. *Occupational Therapy in Health Care*, 32(2), 137-153.

Timmins, N. (2013) *The four UK health systems*. London: The King's Fund.

Touillet, A., Guesdon, H., Bosser, G., Beis, J.M. and Paysant, J. (2010) Assessment of compliance with prescribed activity by hemiplegic stroke patients after an exercise programme and physical activity education. *Annals of Physical and Rehabilitation Medicine*, 53(4), 250-265.

Townsend, E., Stanton, S., Law, M., Polatajko, H., Baptiste, S., Thompson-Franson, T., Kramer, C., Swedlove, F., Brintnell, S. and Campanile, L. (2002) *Enabling occupation: an occupational therapy perspective.* Ottawa, ON: CAOT Publications ACE.

Townsend, E. and Polatajko, H. (2007) *Enabling occupation II: advancing an occupational therapy vision for health, well-being and justice through occupation.* Ottawa, ON: CAOT Publications ACE.

Triandafilou, K.M., Ochoa, J., Kang, X., Fischer, H.C., Ellen Stoykov, M. and Kamper, D.G. (2011) Transient impact of prolonged versus repetitive stretch on hand motor control in chronic stroke. *Topics in Stroke Rehabilitation*, 18(4), 316-324.

Triandafilou, K.M. and Kamper, D.G. (2014a) Carryover effects of cyclical stretching of the digits on hand function in stroke survivors. *Archives of Physical Medicine and Rehabilitation*, 95(8), 1571-1576.

Triandafilou, K.M., Ochoa, J.M. and Kamper, D.G. (2014b) Effect of static versus cyclical stretch on hand motor control in subacute stroke. *International Journal of Neurorehabilitation*, 1(124), 2.

Tuntland, H., Aaslund, M.K., Langeland, E., Espehaug, B. and Kjeken, I. (2016) Psychometric properties of the Canadian Occupational Performance Measure in home-dwelling older adults. *Journal of Multidisciplinary Healthcare*, 9, 411-423.

Turtle, B., Porter-Armstrong, A. and Stinson, M. (2019) A systematic review of the application and psychometric properties of the graded Wolf Motor Function Test. *British Journal of Occupational Therapy*, 83(5), 285-296.

Turtle, B., Porter-Armstrong, A. and Stinson, M. (2020) The reliability of the graded Wolf Motor Function Test for stroke. *British Journal of Occupational Therapy*, In Press. Available at: https://journals.sagepub.com/doi/pdf/10.1177/0308022620902697?casa_token=-mR9vlS9lsUAAAAA:OaFh3kleOJ2SZs_rv6-sSp5BQm7XUoixaNc94c_Q4wpZkDyH3QrA48IrE9RT5fGNO_tGcgi2wT4A [Accessed 3 May 2020].

Uniform Data System for Medical Rehabilitation. (1997) *Guide for the Uniform Data Set for Medical Rehabilitation (Including the FIM Instrument). Version 5.1*. Buffalo, NY: State University of New York.

UK FIM+FAM Users Group. (2011) *The UK FIM+FAM (Functional Assessment Measure). Version 2.2*. Middlesex: UK FIM+FAM Users Group. Available at: https://www.kcl.ac.uk/cicelysaunders/resources/fimfam-manual-v2.2-sept-2012-print-double-sided.pdf [Accessed 15 March 2020].

Uswatte, G., and Taub, E. (2013) Constraint-Induced Movement therapy: a method for harnessing neuroplasticity to treat motor disorders," *In*: Merzenich, M. Nahum, M. and Van Fleet, T. *Changing Brains - Applying Brain Plasticity to Advance and Recover Human Ability*. Burlington: Elsevier, 379–401.

Uswatte, G., Taub, E., Bowman, M.H., Delgado, A., Bryson, C., Morris, D.M., Mckay, S., Barman, J. and Mark, V.W. (2018) Rehabilitation of stroke patients with plegic hands: Randomized controlled trial of expanded constraint-induced movement therapy. *Restorative Neurology and Neuroscience*, 36(2), 225-244.

Van Hout, B., Janssen, M.F., Feng, Y.S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P. and Pickard, A.S. (2012) Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*, 15(5), 708-715.

Van Teijlingen, E. and Hundley, V. (2002) The importance of pilot studies. *Nursing Standard (through 2013)*, 16(40), 33-36.

Vodicka, E., Kim, K., Devine, E.B., Gnanasakthy, A., Scoggins, J.F. and Patrick, D.L. (2015) Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials. gov (2007–2013). *Contemporary Clinical Trials*, 43, 1-9.

Waddell, K.J., Birkenmeier, R.L., Bland, M.D. and Lang, C.E. (2016) An exploratory analysis of the self-reported goals of individuals with chronic upper-extremity paresis following stroke. *Disability and Rehabilitation*, 38(9), 853-857.

Wagner, E. H. (1998) Chronic disease management: what will it take to improve care for chronic illness? *Effective Clinical Practice*, 1(1), 2-4.

Walker, M.F., Sunnerhagen, K.S. and Fischer, R.J. (2013) Evidence-Based Community Stroke Rehabilitation. *Stroke,* 44(1), 293-297.

Wallace, D., Duncan, P. W., and Lai, S. M. (2002) Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *Journal of Clinical Epidemiology*, 55(9), 922-928.

Ward, M.M., Guthrie, L.C. and Alba, M. (2014) Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *Journal of Clinical Epidemiology*, 67(6), 689-696.

West, T. and Bernhardt, J. (2012) Physical activity in hospitalised stroke patients. *Stroke Research and Treatment*, 2012, 813765.

Whitall, J., Savin, D. N., Harris-Love, M., and Waller, S. M. (2006). Psychometric properties of a modified Wolf Motor Function Test for people with mild and moderate upper-extremity hemiparesis. Archives of Physical Medicine and Rehabilitation, 87(5), 656–660.

Winstein, C.J., Stein, J., Arena, R., Bates, B., Cherney, L.R., Cramer, S.C., Deruyter, F., Eng, J.J., Fischer, B., Harvey, R.L. and Lang, C.E. (2016) Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 47(6), e98-e169.

Williamson, P.R., Altman, D.G., Blazeby, J.M., Clarke, M., Devane, D., Gargon, E. and Tugwell, P. (2012) Developing core outcome sets for clinical trials: issues to consider. *Trials*, 13(1), 132.

Winstein, C.J., Stein, J., Arena, R., Bates, B., Cherney, L.R., Cramer, S.C., Deruyter, F., Eng, J.J., Fischer, B., Harvey, R.L. and Lang, C.E. (2016) Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 47(6), e98-e169.

Wolf, S.L., Lecraw, D.E., Barton, L.A. and Jann, B.B. (1989) Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. *Experimental Neurology*, 104(2), 125-132.

Wolf, S.L., Catlin, P.A., Ellis, M., Archer, A.L., Morgan, B. and Piacentino, A. (2001) Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke. *Stroke*, 32(7), 1635-1639.

World Health Organization. (2001) *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization.

Wressle, E., Samuelsson, K. and Henriksson, C. (1999) Responsiveness of the Swedish version of the Canadian occupational performance measure. *Scandinavian Journal of Occupational Therapy*, 6(2), 84-89.

Yardley, L., Ainsworth, B., Arden-Close, E. and Muller, I. (2015) The person-based approach to enhancing the acceptability and feasibility of interventions. *Pilot and Feasibility Studies*, 1(1), 37.

Yavuzer, G., Selles, R., Sezer, N., Sütbeyaz, S., Bussmann, J.B., Köseoğlu, F., Atay, M.B. and Stam, H.J. (2008) Mirror therapy improves hand function in subacute stroke: a randomized controlled trial. *Archives of Physical Medicine and Rehabilitation,* 89(3), 393-398.

Appendices

Appendix 1. Role of research team members for each study chapter

Table 1. Role of team members for Chapter 2

Member of research team	Researcher role
Beverley Turtle	Development of protocol Conducted search of databases Screened titles and abstracts Assessment of articles for inclusion/exclusion criteria Data extraction Synthesised results Write-up of chapter
Dr Alison Porter-Armstrong	Development of protocol Assessment of articles for inclusion/exclusion criteria Intellectual contribution to write-up of chapter and paper
Dr May Stinson	Development of protocol Assessment of articles for inclusion/exclusion criteria Intellectual contribution to write-up of chapter and paper

Table 2. Role of team members for Chapter 3

Member of research team	Researcher role
Beverley Turtle	Conceptualisation and design of study Conducted outcome assessment using videos Data entry, data checking and data cleaning Analysis of results Write-up of chapter and paper
Dr Alison Porter-Armstrong	Ethical and governance approval for pilot RCT Screened potential participants and completed consent with participants for the pilot RCT Conceptualisation and design of study Intellectual contribution to write-up of chapter and paper
Dr May Stinson	Ethical and governance approval for pilot RCT Screened potential participants and completed consent with participants for the pilot RCT Conceptualisation and design of study Intellectual contribution to write-up of chapter and paper
Nicola Gallagher	Conducted and scheduled outcome assessments Data entry
Patricia McIlwaine	Ethical and governance approval for pilot RCT Preliminary screened potential participants for the pilot RCT Training of staff for outcome assessment and mentorship

Lourene Abbi	Preliminary screened potential participants for the pilot RCT Training of staff for outcome assessment
Fiona Morrow	Screened potential participants and completed consent with participants for the pilot RCT
Dr Ian Bradbury	Advice and review of statistical analyses

Table. 3 Role of team members in Chapter 4

Member of research team	Researcher role
Beverley Turtle	Conceptualisation and design of study Data entry, data checking and data cleaning Analysis of results Write-up of chapter Delivered upper limb treatment to both groups
Dr Alison Porter-Armstrong	Ethical and governance approval for pilot RCT Screened potential participants and completed consent with participants for the pilot RCT Conceptualisation and design of study Intellectual contribution to write-up of chapter
Dr May Stinson	Ethical and governance approval for pilot RCT Screened potential participants and completed consent with participants for the pilot RCT Conceptualisation and design of study Intellectual contribution to write-up of chapter and paper
Nicola Gallagher	Conducted and scheduled outcome assessments Data entry
Patricia McIlwaine	Ethical and governance approval for pilot RCT Preliminary screening of potential participants for the pilot RCT Training of staff for outcome assessment and mentorship Delivered upper limb treatment to both groups
Lourene Abbi	Preliminary screening of potential participants for the pilot RCT Training of staff for outcome assessment Delivered upper limb treatment to both groups
Fiona Morrow	Screened potential participants and completed consent for the pilot RCT
Dr Ian Bradbury	Advice and review of statistical analyses

Table 4. Role of team members in Chapter 5

Member of research team	Researcher role
Beverley Turtle	Conceptualisation and design of study Data entry, data checking and data cleaning Analysis of results Write-up of chapter Delivered upper limb treatment to both groups
Dr Alison Porter-Armstrong	Ethical and governance approval for pilot RCT Screened potential participants and completed consent with participants for the pilot RCT Conceptualisation and design of study Intellectual contribution to write-up of chapter
Dr May Stinson	Ethical and governance approval for pilot RCT Screened potential participants and completed consent with participants for the pilot RCT Conceptualisation and design of study Intellectual contribution to write-up of chapter and paper
Nicola Gallagher	Conducted and scheduled outcome assessments Data entry
Patricia McIlwaine	Ethical and governance approval for pilot RCT Preliminary screening of potential participants for the pilot RCT Training of staff for outcome assessment and mentorship Delivered upper limb treatment to both groups
Lourene Abbi	Preliminary screening of potential participants for the pilot RCT Training of staff for outcome assessment Delivered upper limb treatment to both groups
Fiona Morrow	Screened potential participants and completed consent for the pilot RCT
Dr Ian Bradbury	Advice and review of statistical analyses

Table 5. Role of team members in Chapter 6

Member of research team	Researcher role
Beverley Turtle	Conceptualisation and design of study protocol Ethical and governance approval Recruitment and consent Design of focus group topic guide Carried out qualitative analysis Analysis of results Write-up of chapter Delivered upper limb treatment
Dr Alison Porter-Armstrong	Conceptualisation and design of study protocol Ethical and governance approval Design of focus group topic guide Carried out qualitative analysis Intellectual contribution to write-up of chapter
Dr May Stinson	Conceptualisation and design of study protocol Ethical and governance approval Design of focus group topic guide Intellectual contribution to write-up of chapter and paper
Patricia McIlwaine	Ethical and governance approval Design of focus group topic guide Screened and sent out invitation letters to potential participants Delivered upper limb treatment to both groups
Lourene Abbi	Screened and sent out invitation letters to potential participants Reviewed audio recording and transcript for accuracy Delivered upper limb treatment to both groups
Jenny Trainor	Screened and sent out invitation letters to potential participants
Leona Robinson	Focus group moderator Carried out qualitative analysis
Patient and public involvement	Review of study paperwork and focus group topic guide

Appendix 2.

Published systematic review

Title: A systematic review of the application and psychometric properties of the graded Wolf Motor Function Test

Short title: A systematic review of the graded Wolf Motor Function Test

Authors:

1. Beverley Turtle. Centre for Health and Rehabilitation Technologies, Institute of Nursing and Health Research. Ulster University.
2. Alison Porter-Armstrong. School of Health Sciences, Institute of Nursing and Health Research. Ulster University.
3. May Stinson. School of Health Sciences. Institute of Nursing and Health Research. Ulster University.

Corresponding author: Dr Alison Porter-Armstrong, School of Health Sciences, Institute of Nursing and Health Research, Ulster University, Room 01F120, Shore Road, Newtownabbey, BT37 0QB.

Email: a.porter@ulster.ac.uk

Abstract

Introduction: Adapted from the Wolf Motor Function Test, the graded Wolf Motor Function Test (gWMFT) is an upper limb activity assessment for use following stroke and brain injury. The aim of this systematic review was to identify and appraise evidence where the gWMFT has been used or has undergone psychometric evaluation.

Method: A systematic review of five databases was conducted to identify studies reporting the gWMFT using a keyword search. Intervention and clinical measurement studies were eligible for inclusion. Data quality was assessed using adapted Critical Appraisal Skills Programme questions and the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) Risk of Bias checklist.

Results: Twelve studies, of mostly low quality, were included. Studies included one randomised controlled trial, ten pre-and post-studies and one clinical measurement study. All studies involved participants following stroke. Reliability was the only measurement property assessed in two studies, which were of a ‘doubtful’ and ‘poor’ quality.

Conclusion: Low quality studies impede the ability of clinicians and researchers to best determine the applicability of the gWMFT to patient groups and research contexts. Further exploration of the psychometric properties of the gWMFT is recommended across stroke populations using rigorous design methods.

Key words: stroke, outcome assessment (health care), upper limb, psychometrics.

Introduction

Measurement of upper limb function following stroke is complex; the full potential of upper limb motor control cannot be captured through the use of one outcome measure alone (Lang et al., 2013). Upper limb dysfunction can involve paresis, abnormal movement patterns and somatosensory deficits occurring in varying degrees and patterns across stroke survivors (Lang et al., 2013). Therefore, outcome measures must, at least in part, reflect the relevant individualised deficits, as well as demonstrate the expected changes resulting from upper limb treatment (Lang et al., 2013).

The use of standardised outcome measures is encouraged as part of occupational therapy assessment, with emphasis on those that best represent an individual's performance in everyday activities (College of Occupational Therapists, 2017). Using the International Classification of Functioning, Disability and Health (ICF) as a framework to categorise outcome measures (World Health Organisation, 2001), activity-level outcome measures are often viewed as integral to demonstrating meaningful patient outcomes. These are believed to highlight an individual's ability to complete everyday tasks (Lang et al., 2013), which are aligned with the goals of occupational therapy.

The Wolf Motor Function Test (WMFT) (Taub et al., 2011) and the graded Wolf Motor Function Test (gWMFT) (Constraint-Induced Movement Therapy Research Group, 2002) are activity-level assessments of upper limb function (see Table 1). The WMFT is one of the most frequently reported activity-level outcome measures used in investigations of upper limb interventions following stroke (Bushnell et al., 2015; Santisteban et al., 2016). The WMFT was designed to assess the upper limb function of individuals with hemiplegia following stroke or brain injury (Taub et al., 2011). The WMFT consists of 15 items, where individuals are scored on their speed

and quality of performance, and includes two strength tasks (Taub et al., 2011). Items increase in difficulty from assessing joint-specific movements, through to items requiring the performance of functional tasks. The inclusion of functional tasks has led test authors to surmise this assessment may mirror an individual's functional use of their more affected upper limb in everyday life (Morris et al., 2001).

The WMFT has undergone extensive evaluation of its psychometric properties to support its use in stroke rehabilitation. The WMFT has demonstrated high levels of inter-rater reliability (intraclass correlation coefficients >0.9) and test-retest reliability (Pearson's product moment correlation >0.9) for functional ability and performance time when used with individuals more than 12 months' post-stroke (Morris et al., 2001). Adequate responsiveness was found for the WMFT when used with individuals at least six months' post-stroke (Hsieh et al. 2009). Aspects of validity such as construct, predictive and criterion have been demonstrated with significant corrleations found between the WMFT and commonly used outcome measures, including the Functional Independence Measure, Action Research Arm Test and Fugl-Meyer Assessment (Hsieh et al., 2009; Wolf et al., 2001; Wolf et al., 2005). Validated for use with a chronic stroke population, a multidisciplinary panel recommended the WMFT as an outcome measure for use in intervention studies (Bushnell et al., 2015). In a systematic review of upper limb outcome measures used in stroke research, the WMFT was the second most reported outcome measure across 477 studies and the most commonly reported activity-level outcome measure (Santisteban et al., 2016).

However, the WMFT was designed to capture the upper limb capabilities of those with mild to moderate deficits, and primarily those within a chronic stroke population. In response, the test authors developed the gWMFT (Constraint Induced Movement Therapy Research Group, 2002) to capture the activity of individuals with moderate

to severe upper limb impairment and provide more accurate assessment of individuals in the acute or sub-acute stages of stroke.

Table 1

Description of the gWMFT

The gWMFT consists of 13 items, and progresses hierarchically, in a similar pattern as the WMFT (Constraint Induced Movement Therapy Research Group, 2002). Each item consists of two levels (level A and level B), meaning the item can be adjusted to an individual's level of ability. A manual is available for the gWMFT with detailed instructions on test administration and scoring, promoting standardisation (Constraint Induced Movement Therapy Research Group, 2002).

Scoring the gWMFT

Individuals are scored based on their quality of movement and speed of performance. Quality of movement is scored using the functional ability scale (FAS). The FAS is an eight-point ordinal scale, with scores ranging from zero, representing no active movement, through to seven, representing normal movement. Test authors recommend videotaping the assessment and scoring the FAS at a later time to improve accuracy (Constraint Induced Movement Therapy Research Group, 2002). Scores for performance time and the FAS are determined by the level of item completed. For items completed at Level A individuals must complete the item within 30 seconds and can score between four and seven on the FAS. If an individual is unable to complete a test item within 30 seconds, they are then able to attempt the level B version for that item. An extra 60 seconds is added onto the performance time score for items completed at level B. In addition, individuals are only able to score between zero and three on the FAS. A maximum time of 120 seconds is allowed. The test authors recommend the median time and the mean FAS scores to

be reported as summary scores for each individual (Constraint Induced Movement Therapy Research Group, 2002).

However, in comparison to the WMFT, there remains limited uptake of the gWMFT and limited evaluation of its psychometric properties. Such studies are required to determine the suitability of the outcome measure across patient groups and upper limb interventions, as an essential component of evidence-based practice. A broad approach to scoping the literature was adopted, as no synthesis existed which reported upon the clinical use of the gWMFT. This was deemed necessary in order to determine where the gWMFT is being used and with whom, to guide where investigations of its psychometric properties are necessitated. A systematic literature review encompassing examination of the gWMFT is warranted to build on the evidence base for upper limb assessment following stroke.

The aim of this systematic literature review was to explore how the gWMFT has been utilised and reported in the literature. The objectives were:

4. To identify and evaluate studies where the gWMFT has been used as a primary and/or secondary outcome measure.
5. To summarise how the gWMFT has been reported.
6. To identify and evaluate evidence for the measurement properties of the gWMFT.

Method

A systematic literature review was completed by searching the following electronic databases in October 2018: CINAHL (Cumulative Index to Nursing and Allied Health Literature) (1937 to present), Ovid MEDLINE (1966 to present), AMED (1985 to present), PsycINFO (1872 to present) and Pubmed (1947 to present), for the purpose of locating published research regarding the gWMFT. The literature search was developed and completed by the first author with advice from the specialist

subject librarian. The search strategy was formulated using the following keywords in combination: “graded wolf motor function test” OR “gwmft”. The search strategy used for Ovid MEDLINE is detailed in the Appendix.

This systematic literature review was completed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009). This approach was used to enable a transparent and structured search of the literature.

Inclusion and exclusion criteria

Articles were included if they were in the English language and featured the gWMFT as a primary or secondary outcome measure. Due to the limited review scope, and the focus on the use and application of the gWMFT, any studies regardless of patient population, clinical intervention or methodological design were included. Review articles and those where only an abstract was available were excluded.

Study selection

Search results were transferred to the Refworks reference management programme and duplicates were removed. Titles and abstracts for all retrieved studies were screened and examined for any reference to the gWMFT by the first author. The reference list for each relevant publication was also searched, which led to the retrieval of one additional article. The full-text format of papers were reviewed where the outcome measures were not reported in the abstract, and for further examination of study criteria by the first author. The studies retrieved were independently checked by the remaining two review authors, and eligibility confirmed. Differences in opinion were resolved through discussion between the three review authors. One of the full text papers retrieved involved the grade 5 Wolf Motor Function Test (Bowman et al., 2006) and the consensus decision was made to exclude this article.

Data collection and analysis

Data for all included studies were extracted by the first author and recorded using Excel spreadsheets. The data extracted included participant characteristics (age, gender); time post-stroke; study design; intervention applied; and psychometric properties. Aspects of the gWMFT extracted included version reported; and scoring attributes.

The quality of included studies was assessed using an adapted version of the Critical Appraisal Skills Programme (CASP) for cohort studies (Critical Appraisal Skills Programme, 2018), which includes questions used to assess the quality of studies examining outcome measures (Jerosch-Herold, 2005). This combination was chosen due to the varied type of studies found and the focus of the review on outcome measurement.

The quality assessment was as follows:

9. Did the study address a clearly focused issue?
10. Was the sample recruited in an acceptable way?
11. Is the sample size adequate (is there a power calculation)?
12. Is the instrument described and accurately measured to minimise bias?
13. Are the testers trained in test administration?
14. Have the authors identified and taken into account confounding factors?

Where studies examined the psychometric properties of the gWMFT, the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist was used (Mokkink et al., 2018). Designed originally for health-related patient-reported outcomes, the COSMIN can be applied to observer-reported outcome measures, to aid selection and reporting (Hernaez 2015). The measurement property evaluated was rated ‘good’, ‘fair’, ‘doubtful’, ‘poor’ or ‘not applicable’. The lowest rating achieved determined the methodological quality for each applicable study.

Each study was assessed on these attributes by the first author, and scores were agreed upon by the other two authors.

Results

The search results are summarised in the PRISMA flow diagram in Figure 1 (Moher et al., 2009). Thirty-five articles were identified from database and reference list searching. Following the removal of duplicates and application of the inclusion/exclusion criteria, 12 studies were reviewed.

Figure 1

Characteristics of included studies

The characteristics of the included papers are summarised in Table 2. Most studies were performed in the United States of America (n=9), two were completed in India, and one was completed in Brazil. Eleven of the studies were clinically-based, exploring the application of an intervention and these were mostly of a pre- and post-test design. All interventions were aimed at improving upper limb function following stroke. One study was a psychometric investigation examining the inter-rater reliability and agreement of the Brazilian Portuguese version of the gWMFT (Pereira et al., 2015).

The sample sizes for included studies were low; nine studies consisted of 20 participants or less (Bonifer et al., 2005; Bonifer and Anderson, 2003; Demirtas-Tatlidede et al., 2015; Fischer et al., 2016; Flinn et al., 2009; Iwamuro et al., 2011; Pereira et al., 2015; Triandafilou et al., 2011, 2014). Participants were analysed separately according to time post-stroke in the study by Triandafilou and Kamper (2014); 12 participants were two to six months' post-stroke, and 15 participants were more than six months' post-stroke. The study with the largest sample size was completed by Arya et al. (2012) which consisted of 103 participants.

All studies were completed with individuals following stroke, and more than half of the studies were completed with individuals six months or more post-stroke (Bonifer et al., 2005; Bonifer and Anderson, 2003; Demirtas-Tatlidede et al., 2015; Flinn et al., 2009; Iwamuro et al., 2011; Pereira et al., 2015; Triandafilou et al., 2011). Five of the studies came from the same research team, using a device called the X-Glove to passively stretch the fingers of the more affected hand (Fischer et al., 2016; Iwamuro et al., 2011; Triandafilou et al., 2011, 2014; Triandafilou and Kamper, 2014).

Most studies included additional upper limb outcome measures; these are demonstrated in Table 2. The Fugl-Meyer upper extremity scale was most commonly reported, followed by measures of hand and arm strength.

Table 2

Quality of included studies

The quality of studies was generally low (Table 3). The highest scoring study was a randomised controlled trial which completed a sample size calculation, an intention to treat analysis, and an examination of between-group differences in baseline characteristics (Arya et al., 2012).

Most pre- and post-test study designs did not account for how participants were recruited, nor report how bias was reduced. Studies with low sample sizes, and potentially insufficient statistical power were used to determine treatment effects. Most studies did not describe the gWMFT adequately and did not report how the gWMFT was administered and scored. The third most common reason for allocating low scores was due to lack of clarity regarding how test authors reduced confounding factors, including whether participants and assessors were blinded.

There were two case reports which clearly described the participant's health status and the intervention delivered (Bonifer and Anderson, 2003; Flinn et al., 2009).

Table 3

Quality of studies reporting psychometric evaluation

Two studies assessed reliability of the gWMFT (Bonifer et al., 2005; Pereira et al., 2015), and methodological quality was appraised using the COSMIN Risk of Bias checklist (Table 4). Intra-rater reliability of the 14-item gWMFT was assessed as part of the intervention study by Bonifer et al. (2005) and showed a good level of reliability (Pearson's product moment correlation, $r = 0.96$). Paired raters scored functional ability; a physiotherapist and an occupational therapist viewed recorded videos of participants completing the gWMFT. This study received a rating of 'poor' due to the limited methodological detail reported regarding the time interval between scoring sessions and how a final score was achieved between raters (Table 4). In addtion, the Pearson product moment correlation was used, which is not an advised method of analysis (Mokkink et al., 2018).

An inter-rater reliability study of the Brazilian Portuguese version of the 13-item gWMFT was completed by Pereira et al. (2015). Inter-rater reliability and agreement for functional ability and performance time were reported. An excellent level of inter-rater reliability was found for the FAS (intraclass correlation coefficient = 0.98 [95% confidence interval = 0.92-0.99]) and performance time (intraclass correlation coefficient = 0.99 [95% confidence interval = 0.95-1.00]). An adequate amount of agreement was found for the FAS (limits of agreement = -0.68-0.6). Although not noted by study authors, the limits of agreement for performance time indicated inadequate agreement, with limits of agreement between -0.68 and 16.1 seconds. The mean difference between rater scores was 5.5 seconds.

Two raters independently administered the gWMFT and scored performance time and FAS through direct observation to 10 participants more than six months post-stroke. This study received a rating of 'doubtful,' due to lack of clarity in reporting aspects of the reliability analysis (Table 4).

Table 4

Reporting of the gWMFT

The version of the gWMFT used by the studies is demonstrated in Table 5. Although not reported, Pereira et al. (2015) adapted 13 items from the gWMFT into Brazilian Portuguese, indicating the use of the latest version. Similarly, Iwamuro et al. (2011) did not reference the gWMFT manual in their study. However authors report the gWMFT consisted of 13 items, indicating the use of the 2002 version.

The 14-item version of the gWMFT referenced by Bonifer and Anderson (2003) and Bonifer et al. (2005) had an additional item, ‘drop golf ball or washcloth’ and required participants to stand to complete items 10 to 14. Fischer et al. (2016) cite the study by Bonifer and Anderson (2003), indicating the use of the 14-item version.

In the study by Flinn et al. (2009) it was not clear if a variation of the gWMFT was used. The study authors reported that tasks requiring fine motor control were removed; the number of tasks requiring gross motor function were reduced; and tasks which required pronation and supination were increased. There was insufficient reporting in the studies by Anandabai and Gupta (2013) and Demirtas-Tatlidede et al. (2015) to determine the version used.

Table 5

Minimum level of function

Eight of the studies required research participants to have a minimum level of function as part of their eligibility criteria. A significant level of hand impairment was required by Fischer et al. (2016), Iwamuro et al. (2011), Triandafilou and Kamper (2014) and Triandafilou et al. (2014), which was determined by the Chedoke McMaster Stroke Scale for the hand.

It was unclear what criteria were applied in the study by Anandabai and Gupta (2013) and further exploration was not possible. Arya et al. (2012) used the

Brunnstrom stages of arm recovery and those at stages two to five were eligible for their study. This represents a wide variation in ability from basic limb synergies present at stage two to the emergence of more complex movement patterns at stage five. The Fugl-Meyer upper extremity scale was used by Demirtas-Tatlidede et al. (2015), participants were required to score ≤16, indicating a severe level of impairment.

Measurement of the gWMFT

There were variations in how studies reported the scoring criteria for the gWMFT (Table 5), with many not stating the differences in scoring according to level of item completed (Anandabai and Gupta, 2013; Bonifer et al., 2005; Demirtas-Tatlidede et al., 2015; Fischer et al., 2016; Flinn et al., 2009; Iwamuro et al., 2011). Only studies by Arya et al. (2012), Bonifer and Anderson (2003) and Pereira et al. (2015) detailed how the FAS and performance time were scored according to the level of item completed.

Iwamuro et al. (2011), Triandafilou et al. (2011, 2014) and Triandafilou and Kamper (2014) assessed participants on three tasks of the gWMFT. The three tasks included: lifting a pen, lifting cotton balls and lifting a washcloth. Triandafilou et al. (2014) reported the use of three hand-specific items from the gWMFT, which were not detailed.

Scoring criteria was also adapted. In the study by Triandafilou and Kamper (2014) participants received an additional 60 seconds for not using the appropriate grasp, with a maximum of 120 seconds, and the sum score for the three tasks was reported. Triandafilou et al. (2014) reported each task was completed three times for each assessment period, the averages for each task were then summed and used to summarise each assessment session. Triandafilou et al. (2011) reported that each

task was completed in sets of three, with a maximum time of 60 seconds, and did not report what summary score underwent logarithmic transformation.

In the study by Arya et al. (2012) each participant's performance time score was that of their less affected arm subtracted from the score for their more affected arm. Flinn et al. (2009) used the summation of performance time scores as a summary score.

Discussion

This systematic literature review has provided an overview of how the gWMFT has been reported in intervention studies and includes the limited appraisal of its psychometric properties. Although the gWMFT was designed for individuals following stroke or brain injury all included studies involved stroke survivors only. Assessing the quality criteria of the studies, most were of a poor or dubious quality due to the inconsistent administration and scoring of the gWMFT. Due to minimal investigation, evaluation of the psychometric properties of the gWMFT was not used to guide the review. The current review identified one inter-rater reliability study of the Brazilian Portuguese version of the gWMFT (Pereira et al., 2015) and an intra-rater reliability study of the gWMFT FAS (Bonifer et al., 2005).

Hierarchies of evidence are used within healthcare to aid the interpretation of research studies of varying designs (Evans, 2003). Within studies which evaluate the effectiveness of interventions, systematic reviews and randomised controlled trials are viewed as delivering the highest quality of evidence and interpreted as trusted contributors to evidence-based practice (Evans, 2003).

There was one randomised controlled trial included in the review which examined the effectiveness of meaningful task-specific training using the gWMFT as a secondary outcome measure (Arya et al., 2012) and scored highly across all quality criteria. In the study by Arya et al. (2012) the gWMFT was explained in detail; the

scoring criteria for performance time and descriptors of the ordinal scale used to score functional ability were reported aiding study replication.

The remaining intervention studies included in the review consisted of pre- and post-test designs. Non-randomised controlled trials are viewed as liable to increased bias and as such register lower on the hierarchy of evidence (Higgins et al., 2011). The description of outcome measures included was generally of low quality, with some simply stating the gWMFT was used and no further examination provided (Anandabai and Gupta, 2013; Fischer et al., 2016; Iwamuro et al., 2011). Also, how test authors reduced confounding factors were infrequently reported, with many not reporting whether assessors were blinded.

Case reports are descriptive and provide little insight into the efficacy of a treatment and are rated poorly in the hierarchy of evidence (Evans, 2003). However, the case report by Bonifer and Anderson (2003) provided an in-depth description of the training provided to raters, and of the gWMFT administration and scoring. While this study may not score highly in determining effectiveness, this study scored highly in the current review due to how the gWMFT was described and administered. In contrast, the case study by Flinn et al. (2009) poorly described the gWMFT and inaccurately reported the inter-rater reliability study which was completed by Bonifer et al. (2005).

Psychometric properties of the gWMFT

Reliability was the only measurement property assessed for the gWMFT. While the results reported a high level of inter- and intra-rater reliability, the methodological quality of these studies were of a low standard.

In the study by Bonifer et al. (2005), two raters scored functional ability using the videotapes of participants completing the gWMFT. Although the level of training provided was detailed, it was not clear how the two raters came to a final agreed

score for each participant. This study also scored poorly on the COSMIN checklist for not reporting the time interval between the repeated measurements. Reporting the time interval is integral to determine that a long enough period has elapsed to prevent the raters from remembering their previous scores (Kottner et al., 2011). Finally, this study reported reliability using Pearson's product moment correlation which is not an advised method of analysis, with no further exploration of differences between the two-time points (Mokkink et al., 2018).

The study by Pereira et al. (2015) investigated the Brazilian Portuguese version of the gWMFT, which limits its applicability to the English language version. This study scored 'doubtful' on the COSMIN checklist due to lack of clarity regarding the type of intraclass correlation coefficient completed. How an intraclass correlation coefficient is interpreted relies on the type analysed, with assumptions made regarding the number of raters involved and how raters score participants (Kottner et al., 2011). Pereira et al. (2015) also assessed agreement. Bland and Altman plots were used to determine the limits of agreement for scoring functional ability and performance time. Approximately 95% of the difference in scores between raters will lie between the limits of agreement (Bland and Altman, 1999). Ideally, this should be close to zero, indicating minimal differences. A large degree of measurement error was found for scoring performance time, illustrated through wide limits of agreement in the Bland and Altman plot.

In contrast, Fritz et al. (2009) assessed the minimal detectable change for performance time on the WMFT and found that change of at least 0.7 seconds indicated an improvement in ability. The wide degree of measurement error demonstrated by Pereira et al. (2015) would make it difficult to discern whether a change in a participant's score was the result of actual change in recovery or the

result of error. Continued validation of the gWMFT is necessary to determine its ability to accurately measure and document change in upper limb function.

Application of the gWMFT

Across studies, there was heterogeneity in the version of the gWMFT used, how it was applied and scored. Most of these studies did not provide an adequate description of the complexity involved in delivering and scoring the gWMFT. In comparison to the 12 articles suitable for inclusion in this review, a search of the WMFT using Ovid MEDLINE elicited 384 studies. While increased use of the WMFT is to be expected, there remains a wide disparity in uptake between the two outcome measures. Poor reporting in studies of low quality is likely to play a role in whether clinicians or researchers choose to use the gWMFT.

The gWMFT could provide an appropriate alternative to the WMFT for use in the earlier stages of stroke and with those with a greater degree of impairment. However, included studies did not report floor and ceiling effects which would gauge the sensitivity of the gWMFT to measure severe upper limb deficits accurately. Further studies assessing the reliability, validity, and responsiveness of the gWMFT are required to ascertain its appropriateness across interventions and level of impairment.

Adaptation of the WMFT

Through the preparation of this review, another adaptation of the WMFT was found called the Grade 5 WMFT (Uswatte et al., 2018). This test is comparable to the gWMFT in that each item consists of two levels, with similar scoring criteria. The Grade 5 version consists of 10 items, and the mean log score is reported for performance time. The gWMFT was developed to assess the upper limb motor function of individuals with moderate to severe deficits, while the Grade 5 WMFT was developed to assess the motor function of individuals with severe deficits.

The same research team responsible for the development of the WMFT devised a system for classifying the minimum active range of motion required at each joint of the upper limb, which can determine the appropriate WMFT to use (Uswatte et al., 2018; Uswatte and Taub, 2013). The gWMFT is suitable for individuals classified as grades 3/4. This requires stroke survivors to have a minimum of 45 degrees flexion and extension at the shoulder through to the ability to extend at least two fingers to a maximum of 10 degrees and extend the thumb by at least 10 degrees (Uswatte and Taub, 2013).

None of the studies in this review used this classification system as part of their inclusion criteria, with seven studies using three different standardised outcome measures to determine the minimum level of function required. This level of heterogeneity concerning the level of ability for which the gWMFT is appropriate limits its clinical utility.

Implications for research and practice

In summary, this systematic literature review has the following implications for occupational therapists and researchers to consider:

This review has highlighted that in addition to the WMFT, gWMFT and grade 5 WMFT, there exist two versions of the gWMFT. As a result, clinicians and researchers must accurately report the version used and scoring criteria applied. In addition, multiple versions can lead to difficulties in determining the most appropriate outcome measure to use with various client groups. This review has reported that floor and ceiling effects for the gWMFT have not been investigated. Future research should focus on the applicability of the gWMFT to stroke survivors at different stages in their recovery to ascertain its clinical utility.

While there are studies which have reported the gWMFT to a high standard, these were in the minority. Therefore, future attention should be given to the development

of high-quality research measuring upper limb function in stroke survivors using the gWMFT.

There is a paucity in psychometric analysis of the gWMFT, with reliability only examined in two different versions (13- and 14-item) of the outcome measure. However, these studies have limited applicability due to inadequate reporting and includes the assessment of a Brazilian Portuguese version of the gWMFT. Priority should be given to assessing the measurement properties of the gWMFT using rigorous design methods.

Whilst the manual for the WMFT is freely available online, the manual for the gWMFT is available upon request only from the test authors. Potentially impacting on uptake of the gWMFT.

The gWMFT is a multi-component outcome measure. Aspects of the testing kit can be purchased independently for self-assembly, and a template can be purchased directly from the test authors, with the option for self-assembly. However, this has the potential to lead to variations across testing kits, and errors in formation.

There is complexity in the delivery and scoring of the gWMFT. Training would be required to ensure accurate and consistent delivery of the outcome measure. Developed by a research team based in the United States of America, training internationally may not be available to all. These factors could impact reliability, leading to inconsistent delivery and scoring between therapists. As noted in the study by Pereira et al. (2015) there were prominent disagreements between raters scoring, which could have been mitigated by a standardised training protocol.

To improve reduce measurement error, test authors recommend video recording individuals completing the gWMFT for scoring purposes. This would require additional consent procedures to be in place and ensure compliance with general data protection regulation.

Study limitations

This review was limited by the small number of studies identified, with wide heterogeneity impacting data synthesis, which limited the comparisons that could be made. Quality appraisal of the studies was completed using an amalgamation of two quality appraisal tools (CASP, 2018; Jerosch-Herold 2005). Although not a standardised tool, this was created to cope with the heterogeneity of the included studies.

Conclusion

This review has demonstrated that while the gWMFT has limited uptake, researchers are continuing to use this outcome measure with limited evaluation of its measurement properties. To date, there has been an inter-rater reliability and agreement study of performance time and the FAS for the Brazilian Portuguese version, without similar evaluation completed for the English language version. While the intra-rater reliability study was a step in the right direction, this study lacked rigor and only considered the reliability of the FAS. The gWMFT has the potential to extend the applicability of the WMFT and generate meaningful results concerning the recovery of upper limb function in individuals undergoing stroke rehabilitation. However, further investigation is needed to ascertain its application within different stroke populations.

Key findings

- Lack of high-quality upper limb intervention studies which reported the gWMFT.
- Limited psychometric evaluation of the gWMFT, with only reliability assessed.

What the study has added

There is a need for rigorous assessment of the measurement properties and clinical utility of the gWMFT to determine the clinical and research context where it can be best applied.

Acknowledgements: The review team thanks the Ulster University School of Health Sciences librarian who provided advice on the search strategy.

Research ethics: Ethical approval was not required for this study.
Consent: Informed consent was not relevant as this is a systematic review.

Declaration of Conflicting Interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: This work was completed as part of a PhD studentship and was funded by the Department for the Economy (DfE), Northern Ireland.

Contributorship: Beverley Turtle, Alison Porter-Armstrong and May Stinson contributed to the conceptualisation and design of the study. Beverley Turtle undertook the literature search and development of the spreadsheet information in preparation for data analysis. All authors were involved in data analysis, drafting of the manuscript and in continued revisions of the manuscript.

References

Anandabai J and Gupta M (2013) Effects of bimanual functional practice training versus unimanual functional practice training on functional performance of upper extremity in chronic stroke. Journal of Physiotherapy and Sports Medicine 2(1): 15–30.

Arya KN, Verma R, Garg RK, et al. (2012) Meaningful task-specific training (MTST) for stroke rehabilitation: a randomized controlled trial. Topics in Stroke Rehabilitation 19(3): 193–211.

Bland JM and Altman DG (1999) Measuring agreement in method comparison studies. Statistical Methods in Medical Research 8(2): 135–160.

Bonifer N and Anderson KM (2003) Application of constraint-induced movement therapy for an individual with severe chronic upper-extremity hemiplegia. Physical Therapy 83(4): 384–398.

Bonifer NM, Anderson KM and Arciniegas DB (2005) Constraint-induced movement therapy after stroke: efficacy for patients with minimal upper-extremity motor ability. Archives of Physical Medicine and Rehabilitation 86(9): 1867–1873.

Bushnell C, Bettger J, Cockroft K, et al. (2015) Chronic stroke outcome measures for motor function intervention trials: expert panel recommendations. Circulation: Cardiovascular Quality and Outcomes 8(6 suppl 3): S163–S169.

College of Occupational Therapists (2017) Professional Standards for Occupational Therapy Practice. London: College of Occupational Therapists.

Constraint Induced Movement Therapy Research Group (2002) Manual: Graded Wolf Motor Function Test. Birmingham: University of Alabama and Birmingham Veteran's Administration Centre.

Critical Appraisal Skills Programme (2018) CASP cohort study checklist. Available at: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Diagnostic-Checklist-2018.pdf (accessed 11 August 2018).

Demirtas-Tatlidede A, Alonso-Alonso M, Shetty RP, et al. (2015) Long-term effects of contralesional rTMS in severe stroke: Safety, cortical excitability, and relationship with transcallosal motor fibers. NeuroRehabilitation 36: 51–59.

Evans D (2003) Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. Journal of Clinical Nursing 12(1): 77–84.

Fischer HC, Triandafilou KM, Thielbar KO, et al. (2016) Use of a portable assistive glove to facilitate rehabilitation in stroke survivors with severe hand impairment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 24(3): 344–351.

Flinn NA, Smith JL, Tripp CJ, et al. (2009) Effects of robotic-aided rehabilitation on recovery of upper extremity function in chronic stroke: a single case study. Occupational therapy international 16(3-4): 232–243.

Fritz SL, Blanton S, Uswatte G, et al. (2009) Minimal detectable change scores for the Wolf Motor Function Test. Neurorehabilitation and Neural Repair 23(7): 662–667.

Hernaez, R. (2015) Reliability and agreement studies: a guide for clinical investigators. Gut 64 (7): 1018-1027.

Higgins JPT, Altman DG, Gøtzsche PC, et al. (2011) The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. BMJ 343: d5928.

Hsieh Y, Wu C, Lin K, et al. (2009) Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. Stroke 40(4): 1386–1391.

Iwamuro BT, Fischer HC and Kamper DG (2011) A pilot study to assess use of passive extension bias to facilitate finger movement for repetitive task practice after stroke. Topics in Stroke Rehabilitation 18(4): 308–315.

Jerosch-Herold C (2005) An evidence-based approach to choosing outcome measures: a checklist for the critical appraisal of validity, reliability and responsiveness studies. British Journal of Occupational Therapy 68(8): 347–353.

Kottner J, Audig L, Brorson S, et al. (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. Journal of Clinical Epidemiology 64(1): 96–106.

Lang CE, Bland MD, Bailey RR, et al. (2013) Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. Journal of Hand Therapy 26(2): 104–115.

Moher D, Liberati A, Tetzlaff J, et al. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339: b2535.

Mokkink LB, de Vet HCW, Prinsen CAC, et al. (2018) COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Quality of Life Research 27(5): 1171–1179.

Morris DM, Uswatte G, Crago JE, et al. (2001) The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke. Archives of Physical Medicine and Rehabilitation 82(6): 750–755.

Pereira ND, Vieira L, Pompeu FP, et al. (2015) Translation, cultural adaptation and reliability of the brazilian version of the Graded Wolf Motor Function Test in adults with severe hemiparesis. Fisioterapia em Movimento 28(4): 667–676.

Santisteban L, Térémetz M, Bleton J-P, et al. (2016) Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. PLOS ONE 11(5): e0154792.

Taub E, Morris DM, Crago J, et al. (2011) Wolf Motor Function Test (WMFT) Manual. Birmingham: UAB CI Therapy Research Group.

Triandafilou KM and Kamper DG (2014) Carryover effects of cyclical stretching of the digits on hand function in stroke survivors. Archives of Physical Medicine and Rehabilitation 95(8): 1571–1576.

Triandafilou KM, Ochoa J, Kang X, et al. (2011) Transient impact of prolonged versus repetitive stretch on hand motor control in chronic stroke. Topics in Stroke Rehabilitation 18(4): 316–324.

Triandafilou KM, Ochoa J and Kamper D (2014) Effect of static versus cyclical stretch on hand motor control in subacute stroke. International Journal of Neurorehabilitation 01(03): 1–5.

Uswatte G and Taub E (2013) Constraint-induced movement therapy: a method for harnessing neuroplasticity to treat motor disorders. In: Progress in Brain Research. Elsevier, p. Vol. 207, pp. 379–401.

Uswatte G, Taub E, Bowman MH, et al. (2018) Rehabilitation of stroke patients with plegic hands: randomised controlled trial of expanded constraint-induced movement therapy. Restorative Neurology and Neuroscience 36(2): 225–244.

Wolf SL, Catlin PA, Ellis M, et al. (2001) Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke. Stroke; 32(7): 1635–1639.

Wolf, SL, Thompson, PA, Morris, DM, et al. (2005) The EXCITE trial: attributes of the Wolf Motor Function Test in patients with subacute stroke. *Neurorehabilitation and Neural Repair*; *19*(3): 194-205.

World Health Organisation (2001) *International Classification of Functioning, Disability and Health: ICF.* Geneva: World Health Organisation.

Table 1. Items assessed on the Wolf Motor Function Test, 13-item graded Wolf Motor Function Test and 14-item graded Wolf Motor Function Test

Items	Wolf Motor Function Test	13-item graded Wolf Motor Function Test	14-item graded Wolf Motor Function Test
Raise forearm to table (side)[ab]	✓	✓	✓
Raise forearm from table to box (side)[ab]	✓	✓	✓
Extend elbow (side)[ab]	✓	✓	✓
Extend elbow against 1 lb. weight (side)[ab]	✓	✓	✓
Raise hand to table (front)[ab]	✓	✓	✓
Raise hand to box (front)[ab]	✓	✓	✓
Raise weighted hand to box (front)[a]	✓		
Reach and retrieve 1 lb. weight on table (front)[ab]	✓	✓	✓
Raise can to mouth[a]	✓		
Grasp and lift pencil from table[a]	✓		
Lift paperclip from table[a]	✓		
Stack three checkers on top of one another[a]	✓		
Turn over three cards[a]	✓		
Turn key in lock[a]	✓		
Measure grip strength using dynamometer[a]	✓		
Move foam stick through supination and pronation[b]		✓	✓
Grasp and lift washcloth[b]		✓	✓
Flip light switch[b]		✓	✓*
Grasp and lift pen[b]		✓	✓*
Grasp and lift cotton balls[b]		✓	
Lift weighted basket (3 lb.), place onto raised table (standing)[b]		✓	✓*
Grasp and lift foam triangles[c]			✓*
Drop golf ball or washcloth with forearm supported[c]			✓*

[a] Adapted from permission from Constraint Induced Movement Therapy Research Group. Taub E, Morris DM, Crago J, et al. (2011) Wolf Motor Function Test (WMFT) Manual. Birmingham: UAB CI Therapy Research Group.
[b] Adapted with permission from Constraint Induced Movement Therapy Research Group. Constraint Induced Movement Therapy Research Group (2002) Manual: Graded Wolf Motor Function Test. Birmingham: University of Alabama and Birmingham Veteran's Administration Centre.
[c] Adapted with permission from Constraint Induced Movement Therapy Research Group. Constraint-Induced Movement Therapy Research Group (2000) cited in Bonifer N and Anderson KM (2003) Application of constraint-induced movement therapy for an individual with severe chronic upper-extremity hemiplegia. Physical Therapy 83(4): 384–398.

*Task completed in standing.

Table 2. Characteristics of the studies included in the review

Author (year)	Patient (n)	Country	Time post stroke	Age, years (mean ±SD), Gender	Type of study	Intervention	Additional upper limb assessments included
Anandabai and Gupta (2012)	30	India	3-4 months	Not reported 26 male, 6 female	Pre- and post-study	Bimanual (n=15) and unimanual (n=15) functional practice	Fugl-Meyer Assessment
Arya et al. (2012)	Intervention: 51 Control: 52	India	4-24 weeks	Intervention: 51.67 ±7.96 29 male, 22 female Control: 50.21 ±7.60 33 male, 19 female	RCT	Meaningful task-specific training	Action Research Arm Test Fugl-Meyer Assessment
Bonifer and Anderson (2003)	1	USA	15 years	53 1 female	Case report	Constraint-induced movement therapy	Fugl-Meyer Assessment Motor Activity Log
Bonifer et al. (2005)	20	USA	>12 months	57.5 ±16.6 13 male, 7 female	Pre- and post-study/Psychometric study	Constraint-induced movement therapy	Fugl-Meyer Assessment Motor Activity Log
Demirtas-Tatlidede et al., (2015)	10	USA	>1 year	59.5 ±11 4 male, 6 female	Pre- and post-study.	Contra-lesional repetitive transcranial magnetic stimulation	Fugl-Meyer Assessment Hand strength assessments Modified Ashworth Scale
Fischer et al. (2016)	15	USA	2-6 months	63 ±12 10 male, 3 female	Pre- and post-study	Passive cyclical finger stretching with active-assisted task-	Action Research Arm Test Chedoke Arm and Hand Inventory

						oriented training using an orthotic glove	Fugl-Meyer AssessmentHand strength assessments Motor Activity Log
Flinn et al. (2009)	1	USA	15 months	48 1 female	Case report	Robot-assisted therapy	Active range of motion Fugl-Meyer assessment Motor Activity Log
Iwamuro et al. (2011)	5	USA	≥9 months	54 ±11 4 male, 1 female	Pilot pre- and post-study	Passive finger extension using an orthotic glove	Active range of motion Box and Block Test
Pereira et al. (2015)	10	Brazil	>6 months	53.2 ±11.39 6 male, 4 female	Psychometric study	N/A	N/A
Triandafilou and Kamper (2014)	Sub-acute: 12 Chronic: 15	USA	Subacute: 2-6 months Chronic: >7months	Sub-acute: 53 ±6 6 male, 6 female Chronic: 57 ±8 7 male, 8 female	Pre- and post-study	Passive cyclical finger stretching using an orthotic glove	Box and Block Test Hand strength assessments
Triandafilou et al. (2014)	13	USA	2-6 months post	51 ±12 6 male, 7 female	Pre- and post-study	Static finger stretching/passive finger stretching/rest using an orthotic glove	Hand strength assessments Grip termination time
Triandafilou et al. (2011)	15	USA	≥6 months post	57 ±8 7 male, 8 female	Pre- and post-study	Prolonged and repetitive passive finger stretching using an orthotic glove	Hand strength assessments Grip termination time

Note: SD, standard deviation; RCT, randomised controlled trial.

Table 3. Critical appraisal scoring for each reviewed article

Author (year)	**Address focused issue**	**Acceptable recruitment**	**Adequate sample size**	**Instrument described and accurately measured**	**Testers trained in administration**	**Confounding factors identified and taken into account**
Anandabai and Gupta (2012)	Y	CT	N	N	CT	N
Arya et al. (2012)	Y	Y	Y	CT	Y	Y
Bonifer and Anderson (2003)	Y	N/A	N/A	Y	Y	N/A
Bonifer et al. (2005)	Y	Y	N	Y	Y	Y
Demirtas-Tatlidede et al. (2015)	Y	CT	N	N	CT	N
Fischer et al. (2016)	Y	CT	N	N	CT	CT
Flinn et al. (2009)	Y	N/A	N/A	N	Y	N/A
Iwamuro et al. (2011)	Y	CT	N	N	N	CT
Pereira et al. (2015)	Y	CT	N	Y	CT	CT
Triandafilou and Kamper (2014)	Y	CT	N	N	CT	CT
Triandafilou et al. (2014)	Y	CT	N	N	CT	CT
Triandafilou et al. (2011)	Y	CT	N	N	CT	CT

Note: Y, yes; N, no; CT, cannot tell; N/A, not applicable.

Table 4. COSMIN Risk of Bias checklist to assess the methodological quality of the included reliability studies

Design requirements	**Bonifer and Anderson (2005)**	**Pereira et al. (2015)**
Were patients stable in-between measurements?	F	F
Time interval appropriate?	D	G
Conditions similar for both measurements?	F	F
Reliability analysis		
ICC for continuous scores?	F	F
Kappa for dichotomous, ordinal, or nominal scores?	N/A	N/A
Weighted kappa for ordinal scores?	N/A	N/A
Weighting scheme described for ordinal scores?	N/A	N/A
Any important flaws in design?	P	D
Agreement analysis		
Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated for continuous scores	N/A	G
Percentage (positive and negative) agreement calculated for nominal/ordinal scores	N/A	N/A
Any important flaws in design	N/A	D
Final score	P	D

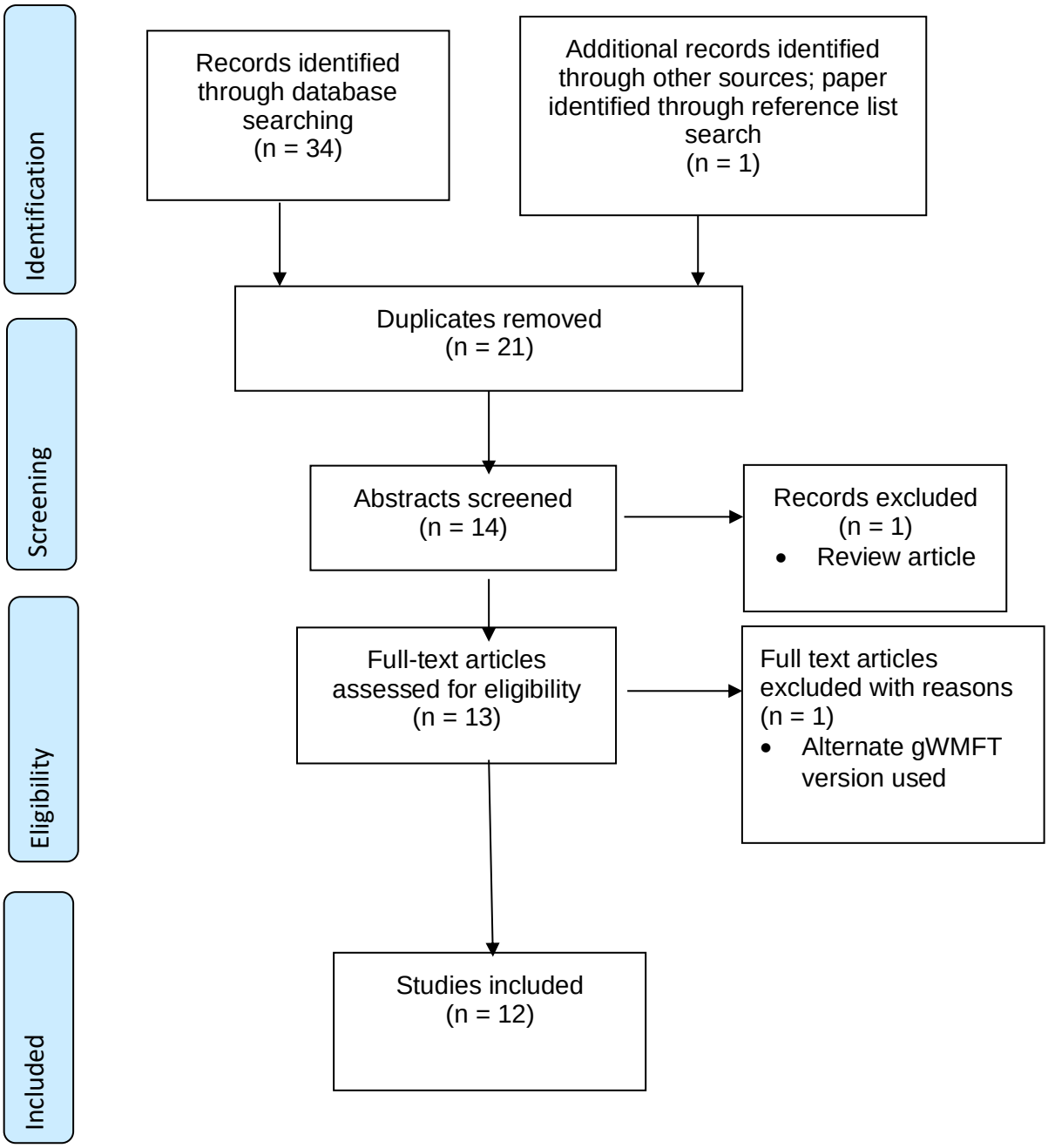
Note: G, good; F, fair; D, doubtful; P, poor; N/A, not applicable

Table 5. Reporting of the gWMFT

Author (year)	13-item/ 14-item gWMFT	Domains of gWMFT scored		Description of scoring criteria
		FAS	Performance time	
Anandabai and Gupta (2012)	Not reported	✓	✓	Not reported
Arya et al. (2012)	13-item	✓	✓	✓
Bonifer and Anderson (2003)	14-item	✓	✓	✓
Bonifer et al. (2005)	14-item	✓	✓	Not reported
Demirtas-Tatlidede et al. (2015)	Not reported	Not reported	Not reported	Not reported
Fischer et al. (2016)	Not clear, cited Bonifer and Anderson (2003)	✓	✓	Not reported
Flinn et al. (2009)	Not clear		✓	Not reported
Iwamuro et al. (2011)	Not clear, reported 13 items		✓	Not reported
Pereira et al. (2015)	13-item	✓	✓	✓
Triandafilou and Kamper (2014)	14-item		✓	✓*
Triandafilou et al. (2014)	14-item		✓	✓*
Triandafilou et al. (2011)	14-item		✓	✓*

Note: ✓, reported; ✓* reported with adaptations

Figure 1. Summary of the literature review search using the PRISMA group flow chart (Moher et al., 2009)



Appendix 3.

List of gWMFT test items and graded options [a]

	Task	Graded options
1	Raise forearm to table (side)	*Level A:* No cushion. *Level B:* Addition of 2.5cm cushion on seat.
2	Raise forearm from table to box (side)	*Level A:* Box at shoulder height. *Level B:* Box at half of shoulder height.
3	Extend elbow (side)	*Level A:* Extend hand to 40cm line. *Level B:* Extend hand to 28cm line.
4	Extend elbow against 1 lb. weight (side)	*Level A:* Extend weight to 40cm line. *Level B:* Extend weight to 28cm line.
5	Raise hand to table (front)	*Level A:* No cushion. *Level B:* Addition of 2.5cm cushion on seat.
6	Raise hand to box (front)	*Level A:* Box at shoulder height. *Level B:* Box at half of shoulder height.
7	Reach and retrieve 1 lb. weight on table	*Level A:* Starting point beyond 40cm line. *Level B:* Starting point beyond 28cm line.
8	Move foam stick through supination and pronation	*Level A:* Participant moves foam stick through supination, touching a box at 5cm, and pronation, touching a box at 2.5cm. *Level B:* Participant moves foam stick through pronation only.
9	Grasp and lift washcloth	*Level A:* Raking grasp is used. *Level B:* Alternate grasp is used.
10	Flip light switch	*Level A:* Lateral pinch grasp is used. *Level B:* Alternate grasp is used.
11	Grasp and lift pen	*Level A:* Tripod grasp is used. *Level B:* Alternate grasp is used.
12	Grasp and lift cotton balls	*Level A:* Tripod grasp is used. *Level B:* Alternate grasp is used.
13	Lift weighted basket (3 lb.), place onto raised table (standing)	*Level A:* Raised table at 22cm above desk. *Level B:* Raised desk lowered to rest upon desk.

[α] Adapted with permission from Constraint Induced Movement Therapy Research Group (2002)

Appendix 4.

Table 1. Scoring procedure for Level A and Level B items of the graded Wolf Motor Function Test

	Performance Time	Functional Ability Scale	
Level A	Score = actual time taken in seconds (0-30 seconds)	7	Task completed. Normal movement
		6	Task completed. Reduced precision, consistency.
		5	Task completed. Noted compensatory movements, increased effort and/or time taken to complete.
		4	Task completed. Slight adjustments made by less affected arm, more than two attempts and/or completed very slowly.
Level B	Score = actual time taken in seconds (0-60 seconds) PLUS additional 60 seconds as Level B tariff	3	Task completed. Noted compensatory movements, increased effort and/or time taken to complete.
		2	Task completed. Slight adjustments made by less affected upper limb, more than two attempts and/or completed very slowly.
		1	No functional movement from more affected upper limb
		0	Unable to complete. No active movement.

Note. Adapted with permission from Constraint Induced Movement Therapy Research Group (2002)

Appendix 5.

MEDLINE Search Strategy

Database: Ovid MEDLINE <1946 to current> Search Strategy:	
1	graded wolf motor function test.mp.
2	gwmft.mp.
3	1 or 2

Appendix 6.

Published Reliability and Agreement study
Title: The reliability of the graded Wolf Motor Function Test for stroke

Short title: Reliability of the gWMFT for stroke

Authors:

4. Beverley Turtle. Centre for Health and Rehabilitation Technologies, Institute of Nursing and Health Research. Ulster University. Newtownabbey, UK.
5. Alison Porter-Armstrong. School of Health Sciences, Institute of Nursing and Health Research. Ulster University. Newtownabbey, UK.
6. May Stinson. School of Health Sciences. Institute of Nursing and Health Research. Ulster University. Newtownabbey, UK.

Corresponding author: Dr Alison Porter-Armstrong, School of Health Sciences, Institute of Nursing and Health Research, Ulster University, Room 01F120, Shore Road, Newtownabbey, BT37 0QB.

Email: a.porter@ulster.ac.uk

Abstract

Introduction: The graded Wolf Motor Function Test assesses upper limb function following stroke. Clinical utility is limited by the requirement to video record for scoring purposes. This study aimed to (1) assess whether video recording is required through examination of inter-rater reliability and agreement; and (2) assess intra-rater reliability and agreement.

Method: A convenience sample of 30 individuals were recruited following stroke. The graded Wolf Motor Function Test was administered within two weeks of rehabilitation commencement and at three months. Two occupational therapists scored participants through either direct observation or video. Inter- and intra-rater reliability and agreement were examined for item-level and summary scores.

Results: Excellent inter-rater reliability (n=28) was found between scoring through direct observation and by video (intraclass correlation coefficients >0.9) and excellent intra-rater reliability (n=21) was found (intraclass correlation coefficients >0.9), for item-level and summary scores. Low agreement was found between raters at item level. Adequate agreement was found for total functional ability, with increased measurement error found for total performance time.

Conclusion: The graded Wolf Motor Function Test is a reliable measure of upper limb function. Video recording may not be required by therapists. In view of low agreement, future studies should assess the impact of standardised training.

Key words: upper limb, outcome assessment, stroke, reliability.

Introduction

Upper limb impairment is common following stroke (Lawrence et al., 2001), with survivors generally experiencing a combination of reduced motor control, coordination and somatosensory deficits (Lang et al., 2013). With links to increased dependence in daily life activities (Lang et al., 2013), improvement in upper limb motor control and function is central to stroke rehabilitation (Pollock et al., 2014). Choice of outcome measure has been identified as one of the top three research priorities for improving clinical trials (Smith et al., 2014). Currently various upper limb outcome measures are recommended according to treatment modality (Sivan et al., 2011), sample group or setting (Langhorne et al., 2011), with no consensus demonstrated in the guidelines (Intercollegiate Stroke Working Party, 2016). The use of standardised outcome measures is essential for evidence-based occupational therapy practice and promoted across occupational therapy guidelines (Association of Canadian Occupational Therapy Regulatory Organizations, 2011; College of Occupational Therapists, 2017; Occupational Therapy Australia, 2018).

The Wolf Motor Function Test (WMFT) was developed to measure upper limb motor activity following stroke and traumatic brain injury (Wolf et al., 1989). Demonstrating adequate psychometric properties among people who have had a stroke (Lin et al., 2009; Morris et al., 2001; Wolf et al., 2001), the WMFT has become a widely used and recommended assessment of upper limb activity (Murphy et al., 2015; Santisteban et al., 2016). The WMFT is recommended for individuals with mild to moderate upper limb impairment (Taub et al., 2011) and is most sensitive to those with a higher level of motor function (Thompson-Butel et al., 2014; Wolf et al., 2001), with floor effects found when used in the early stages of stroke (Lin et al., 2009). The graded Wolf Motor Function Test (gWMFT) was developed for accurate assessment of moderate to severe upper limb impairment (Constraint Induced Movement

Therapy Research Group, 2002). The WMFT and gWMFT are conducted in real time with performances video recorded to reduce measurement error when scoring this complex assessment (Constraint Induced Movement Therapy Research Group, 2002; Taub et al., 2011).

A systematic review explored the clinical application and psychometric properties of the gWMFT reported in the literature (Turtle et al., 2019). This review found that the gWMFT was a secondary outcome measure in 11 clinical trials, with two versions of the outcome measure reported; the 14-item gWMFT, and the more recent, 13-item gWMFT. The studies included in the review were predominantly of low quality due to inconsistencies in how the gWMFT was administered and scored, with some authors adapting it to meet study objectives (Bonifer et al., 2005; Iwamuro et al., 2011; Triandafilou and Kamper, 2014).

Reliability of the two versions of the gWMFT has been assessed across two studies. The 14-item gWMFT was assessed by Bonifer et al. (2005), who found a high level of intra-rater reliability for scoring functional ability in 20 individuals more than 12 months post-stroke. Pereira et al. (2015) found a high level of inter-rater reliability for scoring functional ability and performance time using a Brazilian Portuguese version of the 13-item gWMFT in 10 individuals in the chronic stage of stroke. With no further psychometric evaluation of the gWMFT reported, the gWMFT has limited utility in clinical practice and research. For a more detailed review of the application and psychometric properties of the graded Wolf Motor Function Test, see Turtle et al. (2019).

As noted previously, authors of the gWMFT recommend the use of video recording for scoring participants (Constraint Induced Movement Therapy Research Group, 2002). However, this adds to the burden of delivery and may not be appropriate for

use in clinical practice, with evidence suggesting video recording the WMFT is not required for accurate scoring (Whitall et al., 2006).

Therefore, the aims of the current study were to investigate inter- and intra-rater reliability and agreement, and internal consistency for the gWMFT in a sub-acute stroke population (within three months of stroke onset).

Method

This study is presented based on the published guidelines for reporting reliability and agreement (Kottner et al., 2011). Ethical approval was granted by the Office for Research and Ethics Committees (Ref:14/NI/1149). All participants provided written informed consent.

Participants

Thirty individuals in the sub-acute phase of stroke recruited to an ongoing pilot randomised controlled trial formed the sample (ClinicalTrials.gov: NCT02276729). Inclusion criteria were: adults aged 18 years plus and recently admitted to an inpatient rehabilitation ward; stroke diagnosis within three months with upper limb motor loss and upper limb rehabilitation a key component of treatment; able to understand and follow two-part verbal and written commands in the English language and able to provide written consent. Exclusion criteria were: having had a previous stroke or gross cognitive impairment.

Raters

Rater one and rater two were research occupational therapists. The therapists were employed solely to collect outcome measures on the trial and had no clinical relationship with the participants. Training for both raters involved reviewing the manual (Constraint Induced Movement Therapy Research Group, 2002) and viewing training videos, the scoring of which was verified by occupational therapists experienced in the clinical administration of the outcome.

Outcome measure

The gWMFT assesses timed performance and quality of movement (Constraint Induced Movement Therapy Research Group, 2002). The gWMFT consists of 13 graded test items (Appendix 1) (Constraint Induced Movement Therapy Research Group, 2002) and takes approximately 40 minutes to administer. Video recording of the gWMFT is recommended to enable retrospective scoring of functional ability. A template can be purchased from test authors to standardise placement of the 13 test items.

Video recording

Test items one to eight require placement of the video camera to the side of the template, three feet to the side of the participant being tested, allowing the view of their entire torso (Constraint Induced Movement Therapy Research Group 2002). Test items nine to 12 require the same placement of the video camera but zoomed in to detail the upper limb and fine finger movements. Test item 13 requires placement of the video camera to the front of the template and three feet in front of the participant (Constraint Induced Movement Therapy Research Group 2002).

Scoring of the gWMFT

Quality of movement is assessed on the gWMFT using a functional ability scale (FAS). This is an eight-point ordinal scale, ranging from zero (not attempted) to seven (normal movement). Items are completed on two levels (A and B), where level A items are of a higher level of difficulty and are scored between four and seven. Level B items are of a lower level of difficulty and are scored between zero and three. Any item not completed are scored zero. For the assessment of performance time, participants have 30 seconds to complete level A items, and if unable to do so have a second opportunity to complete the task at level B. Sixty seconds are added

onto performance time for level B items, with a maximum time of 120 seconds. Table 1 presents the scoring procedure for level A and level B test items.

Procedure

The test was administered and video recorded according to protocol guidelines by one occupational therapist (rater one) (Constraint Induced Movement Therapy Research Group, 2002). To standardise placement of objects and participants, the template was devised from a plexiglass sheet according to protocol instructions and securely affixed to a table top (Appendix 2). The gWMFT was used to assess the participants affected arm.

Assessments were completed at two weeks (T1) and three months (T2). The assessments completed at T1 took place in a private room used for research purposes on the hospital site. Assessments completed at T2 generally took place in the participant's own home.

For inter-rater analyses, rater one completed scoring through direct observation and rater two later viewed and scored participant videos for assessments completed at T1.

For intra-rater analyses, rater two scored assessment videos completed at T2 and re-scored one month later.

Internal consistency was assessed using rater two scoring at T1 and T2.

All recorded participant footage was viewed in a private room on hospital premises. Raters were blinded to each other's scoring.

Measurement constructs

Reliability and agreement determine the amount of measurement error in an outcome, and contribute to test validity (Kottner et al., 2011; Streiner et al., 2015). Reliability refers to the amount of variability between rater scores, while agreement assesses the degree to which allocated scores are identical (Kottner et al., 2011;

Streiner et al., 2015). Internal consistency is a form of reliability which assesses the degree to which test items are inter-related and therefore indicative of measuring the same construct (Cronbach, 1951).

Data Analysis

Descriptive statistics for age, gender and side of hemiparesis were recorded. The mean value was reported for the total FAS score, and the median value was reported for total performance time (Constraint Induced Movement Therapy Research Group, 2002). Score distributions were examined for both time points. Floor and ceiling effects were present if 15% or more of the sample achieved the minimum or maximum scores (McHorney and Tarlov, 1995).

Item-level reliability and agreement were completed to determine if there were any issues with individual items of the gWMFT. Inter-rater reliability for total and item-level functional ability and performance time were assessed using a two-way random, consistency intraclass correlation coefficient ($ICC_{2,1}$) (Shrout and Fleiss, 1979). This enables generalisations to be made to other raters within the same population.

Intra-rater reliability for total and item-level functional ability and performance time were assessed using two-way mixed effects, consistency ICC ($ICC_{3,1}$) (Shrout and Fleiss, 1979). Intraclass correlation coefficients determine the level of consistency in the ranking of scores (Hallgren, 2012). A reliability score of 0.60 and above was considered acceptable (Cicchetti, 1994).

To examine item-level inter- and intra-rater agreement, proportion of agreement and proportion of agreement ± 1 point were completed for functional ability. Standard error of measurement (SEM) (Stratford and Goldsmith, 1997) was completed for item-level performance time. Standard error of measurement was calculated for the total scores of both functional ability and performance time. The SEM portrays the

amount of measurement error in scoring; the larger the value, the greater the variability between raters.

Internal consistency of functional ability and performance time were analysed using Cronbach's alpha. Values above 0.70 were considered indicative of test items measuring the same construct and correlating well together (Terwee et al., 2007). All analyses were completed using SPSS Statistics (Version 24.0. IBM Corporation, Armonk, NY).

Results

A total of 30 participants were recruited (mean days post-stroke [SD], 14.73 [8.36]). Due to medical reasons, loss to follow-up and technical difficulties in viewing recorded videos, two and nine participants were not assessed at T1 and T2 respectively. Consequently, data from 28 participants yielded the analyses for inter-rater analyses (mean age [SD], 71.3 [9.85]; 18 males and 10 females) and data from 21 participants yielded the analyses for intra-rater analyses (mean age [SD], 70.5 [8.7]; 16 males and five females).

Technical difficulties prevented the scoring of one item for participant one and one item for participant two at T2. In order to utilise existing data, summary scores were calculated using the available items. Patient characteristics are presented in Table 2.

[Insert Table 2]

Floor and ceiling effects

Ceiling effects were not evident for either assessment session. At T1, floor effects were found for performance time and functional ability by both raters, with 35.7% and 21.4% of the sample achieving the maximum score of 120 seconds and minimum score of zero, respectively (Table 2).

At T2, floor effects were found for performance time, with 33.7% of the sample achieving the maximum score of 120 seconds (Table 2). Floor effects were also

found for functional ability at both testing sessions, with 19% of the sample achieving the minimum score of zero (Table 2).

Inter-rater reliability and agreement

High levels of reliability were found between rater one scoring through direct observation and rater two scoring using recorded videos for item-level (Table 3) and total (Table 4) functional ability and performance time, with ICC values above 0.8. The proportion of agreement for scoring functional ability at item-level ranged from 0.43 to 0.64 and proportion of agreement ± 1 ranged from 0.56 to 0.96 (Table 3). Agreement based on SEM values for performance time at item-level ranged from 0.32 to 19.30, with greater differences found for scoring items one and four through to twelve (Table 3). Standard error of measurement values for total scores were 0.33 for functional ability, and 6.49 for performance time (Table 4). Larger differences for scoring performance time occurred where there were differences between raters in assigning participant performance to level A or level B tasks.

[Insert Table 3]

Intra-rater reliability and agreement

High levels of reliability were found for item-level (Table 3) and total (Table 4) functional ability and performance time, with ICCs above 0.9. Proportion of agreement ranged from 0.57 to 0.86 and proportion of agreement ± 1 ranged from 0.90 to 1 for functional ability scores at item-level (Table 3). Agreement based on SEM values for item-level performance time ranged from 0.07 to 9.29, with greater differences found for scoring items three, four, five, nine, eleven and twelve (Table 3). Standard error of measurement values for total scores were 0.19 for functional ability, and 3.64 for performance time (Table 4).

Internal Consistency

Internal consistency values for functional ability and performance time for both assessment points were above 0.9 (Table 4).

[Insert Table 4]

Discussion

This study estimated the psychometric properties of the gWMFT in a cohort of individuals with stroke and compared the results between scoring through direct observation and using video. Excellent inter-rater reliability was found for the FAS and performance time and adequate agreement was found for scoring functional ability through direct observation and by video. However, unacceptable measurement error was found for scoring performance time. Excellent reliability was also found for intra-rater analyses. This is the first reported study to investigate the reliability and agreement properties of the gWMFT in the sub-acute phase of stroke. With limited psychometric evaluation existing the ability to compare this study to previous literature is limited.

Substantial floor effects were found for performance time, with a high proportion of scores clustering at the maximum performance time allowed. Floor effects for the FAS were found by both raters at T1, and at both testing sessions at T2. Comparable findings were found for the WMFT when used with lower-functioning participants, with five participants unable to complete any item within 120 seconds (Thompson-Butel et al., 2015). Lin et al. (2009) found floor effects for the WMFT FAS when applied within 14 days of stroke onset. A large proportion of the current sample were unable to attempt all test items. With no recorded item available to score, participants scored 120 seconds and zero on the FAS. The pilot study, from which this sample was derived, did not preclude individuals with more severe upper limb impairment from recruitment procedures, potentially explaining the floor effects found. With participants demonstrating varying degrees of upper limb function, the

gWMFT was not able to sensitively measure the range of motor capabilities exhibited.

The high levels of inter-rater reliability found between raters scoring through direct observation and by video indicates that scoring by video may not be a necessary adjunct. This was further substantiated by adequate agreement found between raters for scoring functional ability. While agreement for total FAS scores was adequate, exact agreement was poor across all items. The SEM for performance time highlighted greater discrepancies between raters. Examination of scores at item level highlighted rater variations in assigning participant performance to level A or level B. Examining agreement at item-level, SEM values greater than 9 seconds were found for 10 items. Whilst the raters underwent training separately, the training content was consistent for both. This comprised reading the manual (Constraint Induced movement Therapy Research Group 2002), viewing training videos of an experienced occupational therapist administering the test with stroke survivors and scoring in real time. This was augmented by a review of the scoring results with an experienced occupational therapist in a training session. In previous studies raters have been required to demonstrate approximate scoring to each other prior to study commencement (Morris et al., 2001; Whitall et al., 2006). This was not required in this study, potentially leading to measurement error and the disagreements demonstrated at item-level. Duff et al. (2015) recognised the issues of variability in ascribing the subjective aspects of the WMFT to patient performance and designed a quality process to ensure rater standardisation.

Excellent intra-rater reliability for total and item-level functional ability and performance time were found indicating consistent scoring by one rater, over a one-month interval. Intra-rater SEM values for functional ability displayed minimal variation between scoring sessions, indicating a good level of agreement. Adequate

agreement was found for nine test items, with proportion of agreement greater than 0.7. However, similar to inter-rater agreement analyses, there were unacceptable differences in scoring performance time at both item-level and for total scores.

A previous study has reported good agreement between videotaped and observed scoring for the WMFT based on $ICC_{2,1}$, agreement factor (greater than 0.9) (Whitall et al., 2006). However, the ICC is not a recommended agreement parameter, potentially obscuring the presence of wider variability (Kottner et al., 2011). Whilst differences in scoring modality may have impacted on rater differences in the current study, unacceptable measurement error was found for scoring performance time using video alone. This indicates the presence of additional factors impacting on measurement error. The study authors consider this the result of differences in accurately differentiating between a level A and level B performance by participants. Although, recommended by authors of the gWMFT and the WMFT (Constraint Induced Movement Therapy Research Group, 2002; Taub et al., 2011), the least affected limb was not tested. Scores for the less affected limb may act as a comparison for the more affected limb and help raters discern between FAS ratings accordingly.

Limitations and future research

As part of an ongoing pilot study, the sample size was small, limiting the amount of data available. This study examined participants in the sub-acute phase of stroke, with most experiencing difficulty attempting all test items. Therefore, consideration of reliability and agreement estimates should be applied with caution. Future study could stratify participants according to level of ability and examine use of the gWMFT in chronic stroke. In addition, the grade 5 Wolf Motor Function Test could be used which was developed for individuals with more severe upper limb impairment (Uswatte et al., 2018).

Due to the discrepancies in rater agreement, provision of a standardised training programme throughout may reduce disagreement across level of item assigned, minimising error, and should be considered in future studies.

Implications for Occupational Therapy Practice

The results of this study have the following implications for occupational therapy practice:

- The gWMFT is a reliable measure for assessing upper limb function post-stroke.
- Different therapists could potentially deliver the gWMFT with stroke survivors and score at different time-points, leading to reliable results.
- Given the complexity of the assessment, training would be recommended prior to use, potentially using a fidelity check as developed by Morris et al., (2009) for the WMFT.
- Video recording may not be necessary when scoring the gWMFT, thereby increasing its clinical utility. This would also help to avoid technical errors in video recording and issues with obtaining consent and adhering to General Data Protection Regulations.
- The gWMFT showed floor effects. Therefore, caution should be applied in using the gWMFT with individuals who demonstrate more severe impairments following stroke. The level 5 WMFT could act as a suitable alternative (Uswatte et al., 2018).

Conclusion

The gWMFT demonstrated good levels of inter- and intra-rater reliability and internal consistency. There was acceptable agreement for functional ability, with greater measurement error found for performance time. This study demonstrates the

potential use of the gWMFT in a sub-acute stroke population, without the additive strain of scoring individuals by video.

Key findings:

- The graded Wolf Motor Function Test can be reliably scored by video and/or by direct observation.
- Inadequate agreement for scoring performance time and individual items indicate future studies should consider the impact of standardised training in the use of the assessment.

What the study has added:

The graded Wolf Motor Function Test is a reliable measure of upper limb function in sub-acute stroke and videotaping for scoring purposes may not be required.

Acknowledgement: We wish to thank all participants who took part in the study. We also wish to thank Nicola Gallagher for the acquisition of data and to NICERN for supporting this project. Special thanks to Patricia McIlwaine, Lourene Abbi and Fiona Morrow, occupational therapists at Whiteabbey Hospital, for their assistance with training on outcome administration and scoring, and for their continued support. We would also like to thank Dr Ian Bradbury for his advice and feedback on statistical analyses.

Research ethics: Ethical approval was obtained from the Office for Research Ethics Committees Northern Ireland in 2015 (Ref:14/NI/1149).

Consent: All participants provided written informed consent to participate in the study.

Declaration of conflicting interests: The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: This work was completed as part of a PhD studentship, funded by the Department for the Economy (DfE), Northern Ireland. This study was also supported in part by the United Kingdom Occupational Therapy Research Foundation Research Priority Grant 2014, with additional funding from the Health and Social Care Research Fund, Northern Health and Social Care Trust, Northern Ireland.

Contributorship: All authors contributed to study conception and design. APA and MD applied for ethical approval. BT carried out data collection, statistical analysis and prepared the first draft of this manuscript. All authors were involved in interpretation of the data, reviewed and edited the manuscript, and approved the final version.

References

Association of Canadian Occupational Therapy Regulatory Organizations. (2011). *Essential Competencies of Practice for Occupational Therapists in Canada* (3rd ed.). Toronto: Association of Canadian Occupational Therapy Regulatory Organizations.

Bonifer, N. M., Anderson, K. M., and Arciniegas, D. B. (2005). Constraint-induced movement therapy after stroke: efficacy for patients with minimal upper-extremity motor ability. *Archives of Physical Medicine and Rehabilitation*, *86*(9), 1867–1873. https://doi.org/10.1016/j.apmr.2005.04.002

College of Occupational Therapists (2017) *Professional Standards for Occupational Therapy Practice*. London: College of Occupational Therapists.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Constraint Induced Movement Therapy Research Group. (2002). *Manual: Graded Wolf Motor Function Test*. Birmingham: University of Alabama and Birmingham Veteran's Administration Centre.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Duff, S.V., He, J., Nelsen, M.A., Lane, C.J., Rowe, V.T., Wolf, S.L., Dromerick, A.W. and Winstein, C.J. (2015). Interrater reliability of the Wolf Motor Function Test-Functional Ability Scale: why it matters. *Neurorehabilitation and Neural Repair*, *29*(5), 436–443. https://doi.org/10.1177/1545968314553030

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. 10.20982/tqmp.08.1.p023

Intercollegiate Stroke Working Party. (2016). *National clinical guideline for stroke*.

London: Royal College of Physicians (5th ed.). London: Royal College of Physicians.

Iwamuro, B. T., Fischer, H. C., and Kamper, D. G. (2011). A pilot study to assess use of passive extension bias to facilitate finger movement for repetitive task practice after stroke. *Topics in Stroke Rehabilitation*, *18*(4), 308–315. https://doi.org/10.1310/tsr1804-308

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M. and Streiner, D.L. (2011). Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, *64*(1), 96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002

Lang, C. E., Bland, M. D., Bailey, R. R., Schaefer, S. Y., and Birkenmeier, R. L. (2013). Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. *Journal of Hand Therapy*, *26*(2), 104–115. https://doi.org/10.1016/j.jht.2012.06.005

Langhorne, P., Bernhardt, J., and Kwakkel, G. (2011). Stroke rehabilitation. *The Lancet*, *377*(9778), 1693–1702. https://doi.org/10.1016/S0140-6736(11)60325-5

Lawrence, E. S., Coshall, C., Dundas, R., Stewart, J., Rudd, A. G., Howard, R., and Wolfe, C. D. (2001). Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke*, *32*(6), 1279–1284. https://doi.org/10.1161/01.STR.32.6.1279

Lin, J.-H., Hsu, M.-J., Sheu, C.-F., Wu, T.-S., Lin, R.-T., Chen, C.-H., and Hsieh, C.-L. (2009). Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Physical Therapy*, *89*(8), 840–850. https://doi.org/10.2522/ptj.20080285

McHorney, C. A., and Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research*, *4*(4), 293–307. https://doi.org/10.1007/BF01593882

Morris, D. M., Uswatte, G., Crago, J. E., Cook, E. W., and Taub, E. (2001). The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke. *Archives of Physical Medicine and Rehabilitation*, *82*(6), 750–755. https://doi.org/10.1053/apmr.2001.23183

Morris, D. M., Taub, E., Macrina, D. M., Cook, E. W. and Geiger, B. F. (2009). A method for standardising procedures in rehabilitation: use in the extremity constraint induced therapy evaluation multisite randomised controlled trial. *Archives of Physical Medicine and Rehabilitation*, 90(4), 663-668. https://doi.org/10.1016/j.apmr.2008.09.576

Murphy, M. A., Resteghini, C., Feys, P., and Lamers, I. (2015). An overview of systematic reviews on upper extremity outcome measures after stroke. *BMC Neurology*, *15*(1), 29. https://doi.org/10.1186/s12883-015-0292-6

Occupational Therapy Australia (2018) *Evidence-based practice position statement*. Victoria: Occupational Therapy Australia.

Pereira ND, Vieira L, Pompeu FP, et al. (2015) Translation, cultural adaptation and reliability of the brazilian version of the Graded Wolf Motor Function Test in adults with severe hemiparesis. *Fisioterapia em Movimento* 28(4), 667–676. http://dx.doi.org/10.1590/0103-5150.028.004.AO03

Pollock, A., Farmer, S. E., Brady, M. C., Langhorne, P., Mead, G. E., Mehrholz, J., and van Wijck, F. (2014). Interventions for improving upper limb function after stroke. *The Cochrane Database of Systematic Reviews*, *11*(11), CD010820. https://doi.org/10.1002/14651858.CD010820.pub2

Santisteban, L., Térémetz, M., Bleton, J.-P., Baron, J.-C., Maier, M. A., and Lindberg, P. G. (2016). Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. *PLOS ONE*, *11*(5), e0154792. https://doi.org/10.1371/journal.pone.0154792

Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sivan, M., O'Connor, R., Makower, S., Levesley, M., and Bhakta, B. (2011). Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. *Journal of Rehabilitation Medicine*, *43*(3), 181–189. https://doi.org/10.2340/16501977-0674

Smith, C. T., Hickey, H., Clarke, M., Blazeby, J., and Williamson, P. (2014). The trials methodological research agenda: results from a priority setting exercise. *Trials*, *15*(1), 32. https://doi.org/10.1186/1745-6215-15-32

Stratford, P. W., and Goldsmith, C. H. (1997). Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Physical Therapy*, *77*(7), 745–750. https://doi.org/10.1093/ptj/77.7.745

Streiner, D.L., Norman, G.R. and Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press.

Taub, E., Morris, D.M., Crago, J., King, D., Bowman, M., Bryson, C., Bishop, S., Pearson, S. and Shaw, S.E. (2011). *Wolf Motor Function Test (WMFT) Manual*. Birmingham: UAB CI Therapy Research Group.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M. and de Vet, H.C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Thompson-Butel, A. G., Lin, G. G., Shiner, C. T., and McNulty, P. A. (2014). Two common tests of dexterity can stratify upper limb motor function after stroke. *Neurorehabilitation and Neural Repair*, *28*(8), 788–796. https://doi.org/10.1177/1545968314523678

Thompson-Butel, A. G., Lin, G., Shiner, C. T., and McNulty, P. A. (2015). Comparison of three tools to measure improvements in upper-limb function with poststroke therapy. *Neurorehabilitation and Neural Repair*, 29(4), 341–348. https://doi.org/10.1177/1545968314547766

Triandafilou, K. M., and Kamper, D. G. (2014). Carryover effects of cyclical stretching of the digits on hand function in stroke survivors. *Archives of Physical Medicine and Rehabilitation*, *95*(8), 1571–1576. https://doi.org/http://dx.doi.org/10.1016/j.apmr.2014.04.008

Turtle, B., Porter-Armstrong, A. and Stinson, M. D., (2019) A systematic literature review of the application and psychometric properties of the graded Wolf Motor Function Test. *British Journal of Occupational Therapy*. https://doi.org/10.1177/0308022619879074

Uswatte, G., Taub, E., Bowman, M. H., Delgado, A., Bryson, C., Morris, D. M., McKay, S., Barman, J. and Mark, V. W. (2018). Rehabilitation of stroke patients with plegic hands: randomised controlled trial of expanded constraint-induced movement therapy. *Restorative Neurology and Neuroscience*, 36(2), 225-244. https://doi.org/10.3233/RNN-170792

Whitall, J., Savin, D. N., Harris-Love, M., and Waller, S. M. (2006). Psychometric properties of a modified Wolf Motor Function Test for people with mild and moderate upper-extremity hemiparesis. *Archives of Physical Medicine and Rehabilitation*, *87*(5), 656–660. https://doi.org/10.1016/j.apmr.2006.02.004

Wolf, S. L., Catlin, P. A., Ellis, M., Archer, A. L., Morgan, B., and Piacentino, A. (2001). Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke. *Stroke, 32*(7), 1635–1639. https://doi.org/10.1161/01.STR.32.7.1635

Wolf, S., Lecraw, D., Barton, L., and Jann, B. (1989). Forced use of hemiplegic

upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. *Experimental Neurology*, *104*(2), 125–132. https://doi.org/10.1016/S0014-4886(89)80005-6

Table 1. Scoring procedure for Level A and Level B items of the graded Wolf Motor Function Test

	Performance Time	Functional Ability Scale	
Level A	Score = actual time taken in seconds (0-30 seconds)	7	Task completed. Normal movement
		6	Task completed. Reduced precision, consistency.
		5	Task completed. Noted compensatory movements, increased effort and/or time taken to complete.
		4	Task completed. Slight adjustments made by less affected arm, more than two attempts and/or completed very slowly.
Level B	Score = actual time taken in seconds (0-60 seconds) PLUS additional 60 seconds as Level B tariff	3	Task completed. Noted compensatory movements, increased effort and/or time taken to complete.
		2	Task completed. Slight adjustments made by less affected upper limb, more than two attempts and/or completed very slowly.
		1	No functional movement from more affected upper limb
		0	Unable to complete. No active movement.

Note. Adapted from Constraint Induced Movement Therapy Research Group (2002)

Table 2. Participant characteristics and graded Wolf Motor Function Test scores.

	Two Weeks (T1) (n=28)		Three Months (T2) (n=21)	
Sex				
Male, n	18		16	
Female, n	10		5	
Age in years, mean (SD)	71.3 (9.6)		70.5 (8.7)	
Side of hemiplegia				
Left, n	18		15	
Right, n	10		6	
gWMFT FAS	Rater one	Rater two	Session one	Session two
Mean (SD)	3.74 (2.47)	3.16 (2.11)	3.45 (2.28)	3.53 (2.35)
Floor effect, n (%)	0 (0%)	0 (0%)	4 (19%)	4 (19%)
Ceiling effect, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
gWMFT performance time	Rater one	Rater two	Session one	Session two
Mean (SD)	51.79 (55.18)	53.94 (54.51)	47.74 (55.74)	46.39 (55.77)
Floor effect, n (%)	10 (35.7%)	10 (35.7%)	7 (33.3%)	7 (33.3%)
Ceiling effect, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Abbreviations: gWMFT, graded Wolf Motor Function Test; FAS, functional ability scale; %, percentage

Table 3. Item-level reliability and agreement for the graded Wolf Motor Function Test

	Inter-rater					Intra-rater				
	Reliability $ICC_{(2,1)}$ (95%CI)		Agreement			Reliability $ICC_{(3,1)}$ (95%CI)		Agreement		
	FAS	Time	FAS		Time SEM	FAS	Time	FAS		Time SEM
			Po	Po±1				Po	Po±1	
1- Raise forearm to table (side)	0.884 (0.764-0.944)	0.943 (0.880-0.973)	0.43	0.82	11.11	0.976 (0.940-0.990)	1 (1-1)	0.8	1	0.56
2- Raise forearm from table to box (side)	0.967 (0.930-0.985)	1 (1-1)	0.64	0.93	0.32	0.982 (0.956-0.993)	1 (1-1)	0.81	1	0.54
3- Extend elbow (side)	0.967 (0.931-0.985)	1 (1-1)	0.54	0.96	0.70	0.969 (0.926-0.987)	0.970 (0.929-0.988)	0.71	0.95	9.29
4- Extend elbow against 1 lb. weight (side)	0.866 (0.728-0.937)	0.853 (0.704-0.930)	0.44	0.56	19.30	0.948 (0.875-0.978)	0.967 (0.919-0.986)	0.81	0.95	9.25
5- Raise hand to table (front)	0.913 (0.820-0.959)	0.923 (0.841-0.964)	0.43	0.86	13.22	0.926 (0.827-0.969)	0.969 (0.924-0.987)	0.62	0.95	9.27
6- Raise hand to box (front)	0.926 (0.846-0.965)	0.969 (0.934-0.985)	0.57	0.86	9.89	0.988 (0.970-0.995)	1 (1-1)	0.86	1	0.20
7- Reach and retrieve 1 lb. weight on table	0.953 (0.902-0.978)	0.919 (0.833-0.962)	0.57	0.86	13.65	0.967 (0.920-0.986)	1 (1-1)	0.57	1	0.10
8- Move foam stick through supination and pronation	0.900 (0.797-0.953)	0.901 (0.797-0.953)	0.43	0.93	17.16	0.984 (0.960-0.993)	1 (1-1)	0.76	1	0.07
9- Grasp and lift washcloth	0.946 (0.888-0.975)	0.939 (0.872-0.971)	0.5	0.89	13.43	0.973 (0.936-0.989)	0.972 (0.933-0.989)	0.67	0.90	9.24
10- Flip light switch	0.932 (0.858-0.968)	0.936 (0.867-0.970)	0.57	0.86	13.34	0.976 (0.941-0.990)	1 (1-1)	0.62	1	0.08
11- Grasp and lift pen	0.902 (0.797-0.954)	0.875 (0.745-0.941)	0.52	0.85	17.98	0.954 (0.890-0.981)	0.969 (0.924-0.987)	0.77	0.95	9.24
12- Grasp and lift cotton balls	0.913 (0.820-0.959)	0.914 (0.823-0.959)	0.57	0.82	15.08	0.953 (0.888-0.981)	0.957 (0.898-0.983)	0.86	0.95	9.25

13- Lift weighted basket (3 lb.), place onto raised table (standing)	0.971 (0.939-0.987)	1 (1-1)	0.68	0.93	0.54	0.987 (0.968-0.995)	1 (1-1)	0.8	1	0.08

Abbreviations: ICC, intraclass correlation coefficient; FAS, functional ability scale; Po, proportion of observed agreement; Po±1, proportion of agreement ±1 point; SEM, standard error of measurement

Table 4. Inter- and intra-rater reliability, standard error of measurement and internal consistency of gWMFT.

	Inter-rater reliability	Intra-rater reliability	SEM		Internal Consistency	
	$ICC_{2,1}$ (95% CI)	$ICC_{3,1}$ (95% CI)	Inter-rater	Intra-rater	Two Weeks	Three Months
	(n=28)	(n=21)	(n=28)	(n=21)	(n=28)	(n=19*)
Functional ability	0.979 (0.955-0.990)	0.993 (0.983-0.997)	0.33	0.19	0.99	0.99
Performance time	0.986 (0.970-0.993)	0.996 (0.990-0.998)	6.49	3.64	0.98	0.98

Abbreviations: CI, confidence interval; SEM, standard error of measurement.

*Due to technical difficulties one item was not scored for participants one and two, leading to their exclusion as part of the internal consistency analysis.

Appendix 7.

Ethical approval from the Office for Research Ethics Committee Northern Ireland

HSC Business Services Organisation

Office for Research Ethics Committees Northern Ireland (ORECNI)

Customer Care & Performance Directorate
Office Suite 3
Lisburn Square House
Haslem's Lane
Lisburn
Co. Antrim BT28 1TW
Tel:+44 (0) 28 9260 3107
www.orecni.hscni.net
HSC REC A

07 January 2015

Dr Alison P Porter-Armstrong
Senior Lecturer Rehabilitation Sciences
University of Ulster
Centre for Health and Rehabilitation Technologies
Institute of Nursing and Health Research
School of Health Sciences, Block 1, Jordanstown Campus
BT37 0QB

Dear Dr Porter-Armstrong

Study title:	**A pilot randomized controlled trial (RCT) of mirror box therapy in upper limb rehabilitation with sub-acute stroke patients.**
REC reference:	**14/NI/1149**
IRAS project ID:	**165094**

Thank you for your letter of 05 January 2015, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager, Kathryn Taylor, RECA@hscni.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a **favourable ethical opinion** for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from NRES. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Covering letter on headed paper [cover letter]	1	18 November 2014
Covering letter on headed paper [cover letter addressing Committee concerns]	1	05 January 2015
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Sponsorship Agreement]	1.0	11 November 2014
GP/consultant information sheets or letters [Medical Consultant	1	23 September 2014

Letter]		
IRAS Checklist XML [Checklist_28112014]		28 November 2014
IRAS Checklist XML [Checklist_05012015]		05 January 2015
Letter from sponsor [Sponsorship Letter - Letter from UU confirming co-sponsorship and indemnity arrangements]	1	10 November 2014
Non-validated questionnaire [Exit Questionnaire]	2	05 January 2015
Participant consent form [Consent Form]	1	01 October 2014
Participant consent form [Preliminary Eligibility Screening]	1	23 September 2014
Participant consent form [Eligibility Screen & Consent Process]	1	23 September 2014
Participant consent form [Participant Video Consent form]	1	01 October 2014
Participant consent form [Video Consent Staff Form]	1	01 October 2014
Participant information sheet (PIS) [Participant Information Booklet]	2	05 January 2015
REC Application Form [REC_Form_28112014]		28 November 2014
Referee's report or other scientific critique report [University of Ulster RG3]	1	15 October 2014
Referee's report or other scientific critique report [REC 12/NI/0187 Letter]	1	19 December 2012
Referee's report or other scientific critique report [UKOTRF Panel Review 1 Review]	1	19 March 2014
Referee's report or other scientific critique report [UKOTRF Panel Review Round 2]	1	13 June 2014
Referee's report or other scientific critique report [Prof Avril Drummond Peer Review]	1	04 April 2014
Research protocol or project proposal [protocol]	1	07 October 2014
Summary CV for Chief Investigator (CI) [APA (CI) CV]	1	01 October 2014
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Study Flow Diagram]	1	01 October 2014

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at: http://www.hra.nhs.uk/hra-training/

14/NI/1149	Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely

pp **Dr Catherine Hack**
Chair
Email: RECA@hscni.net

Enclosures: "After ethical review – guidance for researchers"

Copy to: *Mr Nick Curry, University of Ulster*
Frances Johnston, Northern Health and Social Care Trust

Appendix 8.

Governance approval from the Northern Health and Social Care Trust

Northern Health and Social Care Trust

Final Research Governance Permission

7 May 2015

Dr Alison Porter-Armstrong
Senior Lecturer Rehabilitation Sciences
University of Ulster
Centre for Health and Rehabilitation Technologies
Institute of Nursing and Health Research
School of Health & Rehabilitation Technologies
Institute of Nursing Health Research
School of Health Sciences Block 1
Jordanstown Campus
BT37 0QB

Dear Dr Porter-Armstrong

Study Title: Mirror box therapy with stroke patients - A pilot randomized controlled trial (RCT) of mirror box therapy in upper limb rehabilitation with sub-acute stroke patients

HSC Trust Ref: NRP14-0441-11 (Please quote this number in all future correspondence)
REC Ref: 14/NI/1149
IRAS Ref: 165094

I am pleased to advise that the Northern Health & Social Care Trust has given Final Research Governance Permission for the above project to commence. Permission is granted for the duration of the project to 30 June 2017.

The following documents have been approved for use in the project:

Document	Version	Date
Protocol	V1	07-10-2014
Participant Information Sheet (Patient information Booklet)	V2	05-01-2015
Participant Consent Form (Consent Form UK0TRF 2014)		01-10-2014
Participant Consent Form (Preliminary Screening Form)	V1	23-09-2014
Participant Consent Form (Eligibility and Screening Form)	V1	23-09-2014
GP/Consultant information sheets or letters	V1	23-09-2014
Non-validated questionnaire (Upper Limb Questionnaire)	V2	05-01-2015
Letter from Funder (Funding Agreement UU/NHSCT)		
Sponsorship Agreement		03-12-2014
Evidence of Sponsor/Indemnity		10-11-2014
gWMFTVideo Consent Form	V1	Oct 2014
gWMFTVideo Staff Consent Form	V1	Oct 2014
Mrs Patricia McIlwaine CV & GCP		NHSCT
Dr Lourene Abbi - CV & GCP		NHSCT
Miss Fiona Morrow - CV & GCP		NHSCT
Mrs Jennifer Trainor - CV & GCP		NHSCT
Dr May Stinson - CV & GCP		UoU
Professor Ian Bradbury - CV & GCP		UoU
Ms Nicola Gallagher - CV & GCP		UoU
Dr Alison Porter-Armstrong - CV & GCP		UoU

The following personnel have been approved to work on the study at this Trust:

Name	Indemnity Provided by
Mrs Patricia McIlwaine	NHSCT
Dr Lourene Abbi	NHSCT
Miss Fiona Morrow	NHSCT
Mrs Jennifer Trainor	NHSCT
Dr May Stinson	UoU
Professor Ian Bradbury	UoU
Ms Nicola Gallagher	UoU
Dr Alison Porter-Armstrong	UoU

Permission is granted subject to the attached conditions and I would ask you to please ensure that all members of the research team are familiar with these. Failure to abide by these conditions will invalidate permission and may result in the cessation of the research.

I wish you every success with your project.

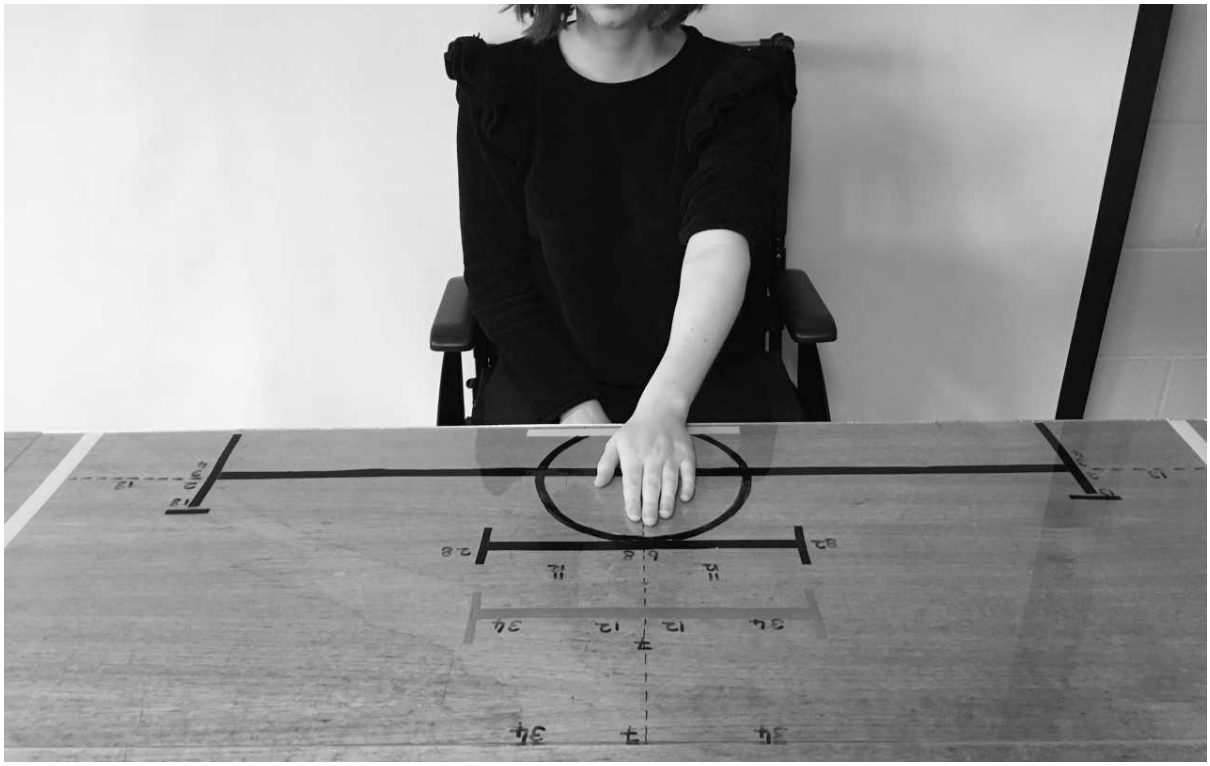
Yours sincerely,

Dr Desmond Rooney
Head of NHSCT R&D

CC Mrs Patricia McIlwaine, NHSCT
Dr Lourene Abbi, NHSCT
Miss Fiona Morrow, NHSCT
Mrs Jennifer Trainor, NHSCT
Dr May Stinson, UoU
Professor Ian Bradbury, UoU
Ms Nicola Gallagher, UoU
Mr Nick Curry, UoU

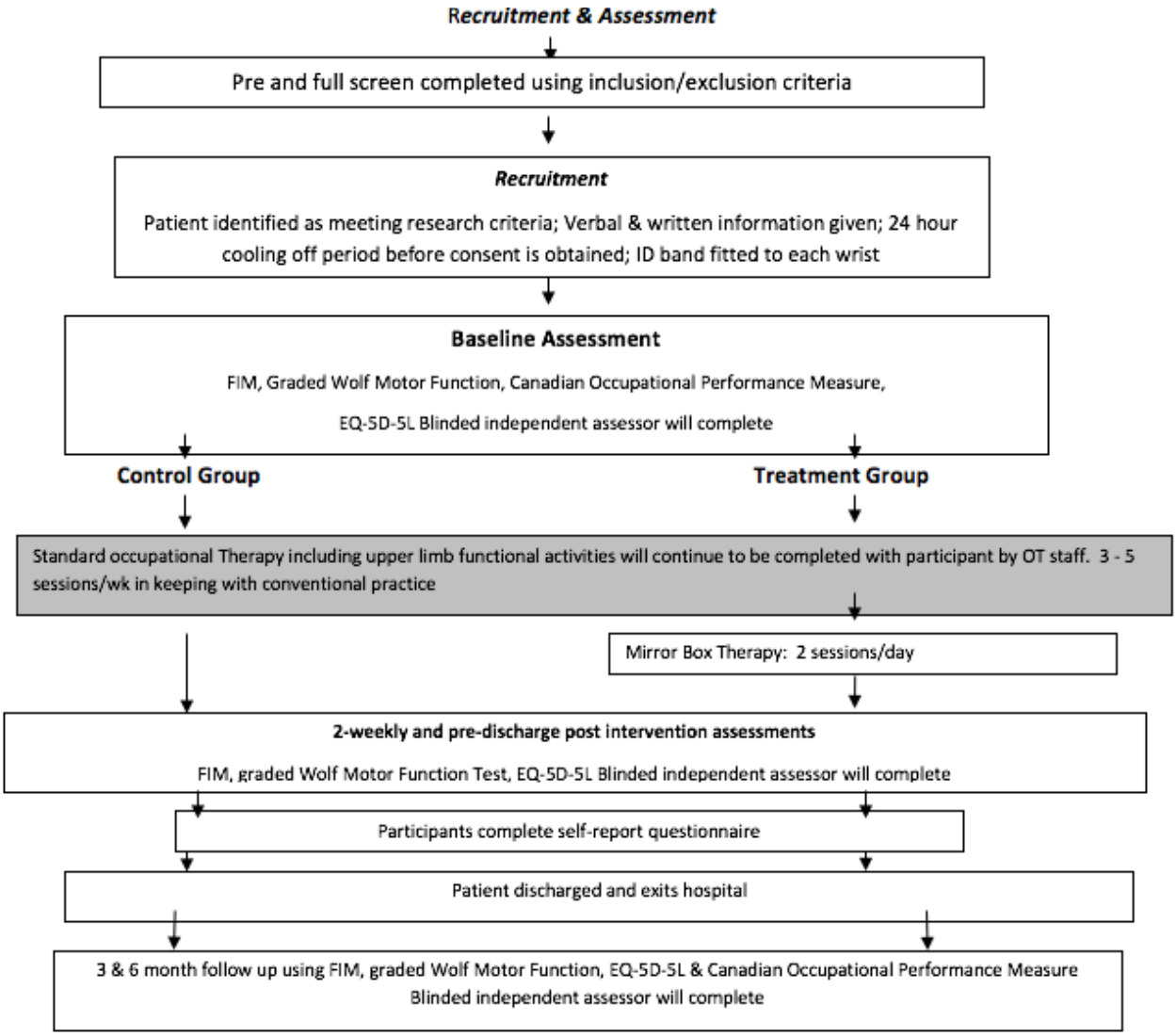
Appendix 9.

Photographic layout of graded Wolf Motor Function Test



Appendix 10.

Flow diagram of study procedure for pilot randomised controlled trial



Appendix 11.

EQ-5D-5L assessment

Health Questionnaire

English version for the UK

Under each heading, please tick the ONE box that best describes your health TODAY.

MOBILITY

I have no problems in walking about ❑
I have slight problems in walking about ❑
I have moderate problems in walking about ❑
I have severe problems in walking about ❑
I am unable to walk about ❑

SELF-CARE

I have no problems washing or dressing myself ❑
I have slight problems washing or dressing myself ❑
I have moderate problems washing or dressing myself ❑
I have severe problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities ❑
I have slight problems doing my usual activities ❑
I have moderate problems doing my usual activities ❑
I have severe problems doing my usual activities ❑
I am unable to do my usual activities ❑

PAIN / DISCOMFORT

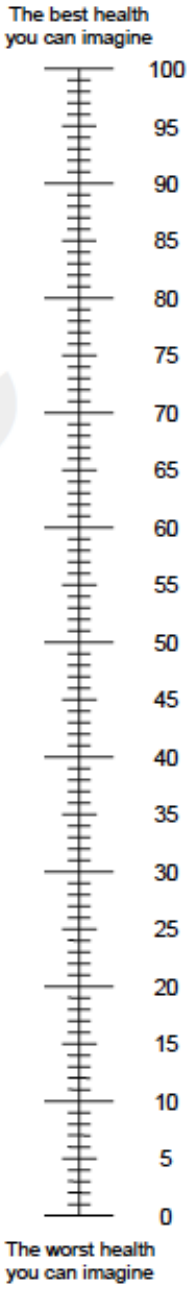
I have no pain or discomfort ❑
I have slight pain or discomfort ❑
I have moderate pain or discomfort ❑
I have severe pain or discomfort ❑
I have extreme pain or discomfort ❑

ANXIETY / DEPRESSION

I am not anxious or depressed ❑
I am slightly anxious or depressed ❑
I am moderately anxious or depressed ❑
I am severely anxious or depressed ❑
I am extremely anxious or depressed ❑

- We would like to know how good or bad your health is TODAY.
- This scale is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Please mark an X on the scale to indicate how your health is TODAY.
- Now, write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐



Appendix 12.

Occupational performance problems identified by the Canadian Occupational Performance Measure

Occupational performance problems chosen by participants

Category		Frequency
Self-care	Getting In/Out Bed	11
	Making Cup Of Tea	1
	Toilet	15
	Shower	5
	Stairs	5
	Driving	17
	Using Knife/Fork	1
	Washed/Dressed	17
	Opening Containers	2
	Shopping	1
	Getting Clothes	1
	Grooming	5
	Walking	21
	Walking Dogs	1
	Mobility Scooter	1
	Public Transport	2
Productivity	Choosing Meals	1
	Use of eBay	1
	Groceries	5
	Caring	5
	Looking After Animals	4
	Housework	8
	Ironing	1
	Laundry	2
	Selling Cattle	1
	Work	5
	Gardening	2
	Community Work	1
	Role as Mother/Grandmother	1
	Cooking	7
	Making the bed	1
	Sewing/Jam	1
Leisure	Going to church	2
	Watching Tv	2
	Kindle	1

	Walk Dogs	4
	Social outings	8
	Walking (Outside)	3
	gardening	4
	golf	2
	social club	2
	reading	4
	listening to radio	1
	hobby	8
	writing	1

Appendix 13.

Frequency distributions for responses to the EQ-5D-5L dimensions

		Baseline (n=39)	Discharge (n=35)	Three-month (n=32)	Six-month (n=25)
Mobility	**No problems**	0 (0%)	4 (11.4%)	3 (9.4%)	1 (4%)
	Slight problems	1 (2.6%)	9 (25.7%)	7 (21.9%)	13 (52%)
	Moderate problems	5 (12.8%)	12 (34.3%)	13 (40.6%)	7 (28%)
	Severe problems	12 (30.8%)	8 (22.9%)	8 (25%)	2 (8%)
	Unable to	21 (53.8%)	2 (5.7%)	1 (3.1%)	2 (8%)
Self-care	**No problems**	3 (7.7%)	6 (17.1%)	9 (28.1%)	7 (28%)
	Slight problems	7 (17.9%)	11 (31.4%)	10 (31.3%)	8 (32%)
	Moderate problems	10 (25.6%)	14 (40%)	9 (28.1%)	7 (28%)
	Severe problems	11 (28.2%)	3 (8.6%)	3 (9.4%)	1 (4%)
	Unable to	8 (20.5%)	1 (2.9%)	1 (3.1%)	2 (8%)
Activity	**No problems**	2 (5.1%)	4 (11.4%)	1 (3.1%)	3 (12%)
	Slight problems	0 (0%)	6 (17.1%)	6 (18.8%)	5 (20%)
	Moderate problems	5 (12.8%)	8 (22.9%)	10 (31.3%)	9 (36%)
	Severe problems	12 (30.8%)	12 (34.3%)	9 (28.1%)	3 (12%)
	Unable to	20 (51.3%)	5 (14.3%)	6 (18.8%)	5 (20%)
Pain	**No pain**	13 (33.3%)	15 (42.9%)	5 (15.6%)	5 (20%)
	Slight pain	8 (20.5%)	12 (34.3%)	11 (34.4%)	10 (40%)
	Moderate pain	15 (38.5%)	7 (20%)	11 (34.4%)	7 (28%)
	Severe pain	2 (5.1%)	1 (2.9%)	3 (9.4%)	2 (8%)

	Extreme pain	1 (2.6%)	0 (0%)	2 (6.3%)	1 (4%)
Anxiety	**Not anxious**	9 (23.1%)	14 (40%)	11 (34.4%)	10 (40%)
	Slightly anxious	11 (28.2%)	11 (31.4%)	13 (40.6%)	11 (44%)
	Moderately anxious	13 (33.3%)	6 (17.1%)	5 (15.6%)	2 (8%)
	Severely anxious	2 (5.1%)	1 (2.9%)	0 (0%)	0 (0%)
	Extremely anxious	4 (10.3%)	3 (8.6%)	3 (9.4%)	2 (8%)

Appendix 14.

Frequency distributions for responses to the EQ-5D-5L dimensions by treatment group

		Baseline		Discharge		Three-month		Six-month	
		Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control
Mobility	No problems	0 (0%)	0 (0%)	3 (16.7%)	1 (5.9%)	3 (16.7%)	0 (0%)	1 (6.7%)	0 (0%)
	Slight problems	1 (5.3%)	0 (0%)	6 (33.3%)	3 (17.6%)	5 (27.8%)	2 (14.3%)	9 (60%)	4 (40%)
	Moderate problems	2 (10.5%)	3 (15%)	4 (22.2%)	8 (47.1%)	6 (33.3%)	7 (50%)	4 (26.7%)	3 (30%)
	Severe problems	5 (26.3%)	7 (35%)	4 (22.2%)	4 (23.5%)	3 (16.7%)	5 (35.7%)	0 (0%)	2 (20%)
	Unable to	11 (57.9%)	10 (50%)	1 (5.6%)	1 (5.9%)	1 (5.6%)	0 (0%)	1 (6.7%)	1 (10%)
Self-care	No problems	2 (10.5%)	1 (5%)	5 (27.8%)	1 (5.9%)	6 (33.3%)	3 (21.4%)	5 (33.3%)	2 (20%)
	Slight problems	3 (15.8%)	4 (20%)	6 (33.3%)	5 (29.4%)	7 (38.9%)	3 (21.4%)	6 (40%)	2 (20%)
	Moderate problems	4 (21.1%)	6 (30%)	6 (33.3%)	8 (47.1%)	4 (22.2%)	5 (35.7%)	3 (20%)	4 (40%)
	Severe problems	7 (36.8%)	4 (20%)	1 (5.6%)	2 (11.8%)	0 (0%)	3 (21.4%)	0 (0%)	1 (10%)
	Unable to	3 (15.8%)	5 (25%)	0 (0%)	1 (5.9%)	1 (5.6%)	0 (0%)	1 (6.7%)	1 (10%)
Activity	No problems	2 (10.5%)	0 (0%)	2 (11.1%)	2 (11.8%)	1 (5.6%)	1 (7.1%)	2 (13.3%)	1 (10%)
	Slight problems	0 (0%)	0 (0%)	5 (27.8%)	1 (5.9%)	5 (27.8%)	0 (0%)	4 (26.7%)	1 (10%)
	Moderate problems	2 (10.5%)	3 (15%)	5 (27.8%)	3 (17.6%)	6 (33.3%)	4 (28.6%)	5 (33.3%)	4 (40%)
	Severe problems	7 (36.8%)	5 (25%)	3 (16.7%)	9 (52.9%)	4 (22.2%)	5 (35.7%)	1 (6.7%)	2 (20%)
	Unable to	8 (42.1%)	12 (60%)	3 (16.7%)	2 (11.8%)	2 (11.1%)	4 (28.6%)	3 (20%)	2 (20%)
Pain	No pain	8 (42.1%)	5 (25%)	10 (55.6%)	5 (29.4%)	4 (22.2%)	1 (7.1%)	5 (33.3%)	0 (0%)
	Slight pain	3 (15.8%)	5 (25%)	6 (33.3%)	6 (35.3%)	5 (27.8%)	6 (42.9%)	7 (46.7%)	3 (30%)

	Moderate pain	6 (31.6%)	9 (45%)	2 (11.1%)	5 (29.4%)	6 (33.3%)	5 (35.7%)	3 (20%)	4 (40%)
	Severe pain	1 (5.3%)	1 (5%)	0 (0%)	1 (5.9%)	2 (11.1%)	1 (7.1%)	0 (0%)	2 (20%)
	Extreme pain	1 (5.3%)	0 (0%)	0 (0%)	0 (0%)	1 (5.6%)	1 (7.1%)	0 (0%)	1 (10%)
Anxiety	**Not anxious**	5 (26.3%)	4 (20%)	9 (50%)	5 (29.4%)	10 (55.6%)	1 (7.1%)	8 (53.3%)	2 (20%)
	Slightly anxious	5 (26.3%)	6 (30%)	7 (38.9%)	4 (23.5%)	6 (33.3%)	7 (50%)	6 (40%)	5 (50%)
	Moderately anxious	6 (31.6%)	7 (35%)	1 (5.6%)	5 (29.4%)	1 (5.6%)	4 (28.6%)	0 (0%)	2 (20%)
	Severely anxious	0 (0%)	2 (10%)	0 (0%)	1 (5.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Extremely anxious	3 (15.8%)	1 (5%)	1 (5.6%)	2 (11.8%)	1 (5.6%)	2 (14.3%)	1 (6.7%)	1 (10%)

Appendix 15.

Positionality statement

Positionality

The PhD researcher is a 34-year-old woman and occupational therapist with two years' experience working in a part-time capacity on the stroke rehabilitation ward where participants received their treatment. Although not the lead therapist for participants in the study, there were occasions where the PhD researcher completed occupational therapy treatments, including mirror therapy with individuals recruited to the study. As such with insider knowledge of treatment received and how it was delivered on the study site, the PhD researcher will hold their own personal views on barriers and facilitators to mirror therapy (Burns *et al*. 2012). By adopting a reflexive approach, assumptions were acknowledged and set aside to ensure prioritisation of participants views and experiences (Underwood *et al*. 2010).

In light of the minimal role the PhD researcher played in day-to-day ward activities and irregular contacts with participants while in hospital, it was considered their presence would have minimal impact on responses. However, to minimise potential bias the focus group was led by an occupational therapist unknown to the participants, with the PhD researcher present as a note taker.

Appendix 16.

Research Ethics Committee approval

Health Research Authority

London - South East Research Ethics Committee
Barlow House
3rd Floor
4 Minshull Street
Manchester
M1 3DZ

Telephone: 0207 104 8002

21 November 2018

Dr Alison Porter-Armstrong
Centre for Health and Rehabilitation Technologies
Institute of Nursing and Health Research
School of Health Sciences, Block 1
Jordanstown Campus
Ulster University
BT37 0QB

Dear Dr Porter-Armstrong

Study title:	**Stroke survivors' views on mirror therapy in upper limb rehabilitation.**
REC reference:	**18/LO/2032**
IRAS project ID:	**249063**

The Proportionate Review Sub-committee of the London - South East Research Ethics Committee reviewed the above application via correspondence. Thank you for being available via email to provide responses to queries from the PRS Sub-Committee.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact hra.studyregistration@nhs.net outlining the reasons for your request. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

Extract of the meeting minutes

Recruitment arrangements and access to health information, and fair participant selection

The PRS Sub-Committee sought further clarification as to who would make and record the clinical assessment that potential participants had no issues with capacity to understand the Information Sheet and Consent Form.

Miss Beverley Turtle replied that a member of the clinical team would verify that potential participants have no issues with capacity to understand the Information Sheet and Consent Form.

The PRS Sub-Committee sought further clarification that potential participants would not have a visual disorder that might interfere with the effectiveness of the mirror therapy.

Miss Turtle replied that potential participants would not have a visual disorder that might interfere with the effectiveness of mirror therapy. She went on to say that this would have been verified by occupational therapists working with the potential participants at the time of their stroke rehabilitation.

The PRS Sub-Committee was satisfied with the responses.

Care and protection of research participants; respect for potential and enrolled participants' welfare and dignity

The PRS Sub-Committee noted that the focus groups would be recorded. They wondered what would happen if some of the participants in the focus groups did not want to be recorded, for instance, the PRS Sub-Committee asked if they would not be able to take part in the group or would there be a focus group specifically for those who are unwilling to be recorded.

Miss Turtle replied that due to limited resources only those able to take part in the recorded focus group sessions would be able to take part in the focus groups.

The PRS Sub-Committee was content with the response.

Informed consent process and the adequacy and completeness of participant information

The PRS Sub-Committee commented that the Information Sheet did not state who the researchers were and that it was part of a PhD project. Therefore, they requested that there should be a statement at the beginning of the Information Sheet detailing who the main researcher is and that they are from Ulster University and that the research would be carried out as part of a PhD in conjunction of the OT department of the NHSC Trust.

Miss Turtle submitted a revised Information Sheet with the requested changes.

The PRS Sub-Committee was content with the revised Information Sheet.

Approved documents

The documents reviewed and approved were:

Document	*Version*	*Date*
Covering letter on headed paper [Ulster university covering letter]	1	16 October 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Sponsorship documents]		19 October 2018
Interview schedules or topic guides for participants [Focus	1	07 August 2018

group topic guide]		
IRAS Application Form [IRAS_Form_30102018]		30 October 2018
IRAS Checklist XML [Checklist_30102018]		30 October 2018
Letter from sponsor [Sponsorship documents]	1	19 October 2018
Letters of invitation to participant [Participant information letter]	3	08 October 2018
Letters of invitation to participant [Invitation letter reminder]	1	24 September 2018
Non-validated questionnaire [Demographic questionnaire]	1	07 August 2018
Other [Response to PR Sub-Committee Queries]		20 November 2018
Participant consent form [Consent form]	2	24 September 2018
Participant information sheet (PIS)	3	20 November 2018
Referee's report or other scientific critique report [Filter committee report - RG3]		05 October 2018
Research protocol or project proposal [qualitative protocol]	3	02 October 2018
Summary CV for Chief Investigator (CI) [Dr Alison Porter-Armstrong CV]		03 September 2018
Summary CV for student [Student CV]		03 September 2018
Summary CV for supervisor (student research) [Supervisor CV]		03 September 2018
Summary of any applicable exclusions to sponsor insurance (non-NHS sponsors only) [Sponsorship documents]		19 October 2018
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Study flowchart]	2	22 August 2018

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

18/LO/2032	Please quote this number on all correspondence

Yours sincerely

On behalf of
Ms Stephanie Cooper
Chair

Email: nrescommittee.london-southeast@nhs.net

Enclosures: List of names and professions of members who took part in the review

"After ethical review – guidance for researchers"

Copy to: Mr Nick Curry

Mrs Frances Johnston,
Northern Health and Social Care Trust

Lead Nation Northern Ireland: research.gateway@hscni.net

NHS
Health Research Authority

London - South East Research Ethics Committee

Attendance at PRS Sub-Committee of the REC meeting

Committee Members:

Name	*Profession*	*Present*	*Notes*
Ms Stephanie Cooper Chair	Retired Solicitor	Yes	
Professor Anthony Fox	Pharmaceutical Medicine	Yes	
Ms Brigid Tucker	Comms Consultant	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Miss Ewa Grzegorska	
Mrs Margaret Hutchinson	REC Manager

Appendix 17.

Governance approval from Northern Health and Social Care Trust

Northern Health and Social Care Trust

www.northerntrust.hscni.net

Governance Department

Final Research Governance Permission

5 February 2019

Dr Alison Porter-Armstrong
Centre for Health & Rehabilitation Technologies
Institute of Nursing and Health Research
School of Health Sciences
Block 1
Jordanstown Campus
Ulster University
BT37 OQB

Dear Dr Porter-Armstrong

Study Title:	Stroke survivor's views on mirror therapy in upper limb rehabilitation
NHSCT Ref:	NT18-0644-12
REC Ref Number:	18/LO/2032
IRAS project ID:	249063

I am pleased to advise that the Northern Health & Social Care Trust has given Final Research Governance Permission for the above project to commence. Permission is granted for the duration of the project to 31 July 2019.

The following documents have been approved for use in the project:

Document	Version	Dated
Covering letter on headed paper (UU Covering letter)		16/10/2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) (Sponsorship documents)		19/10/2018
Interview schedules or topic guides for participants (focus group topic guide)	1	07/08/2018
IRAS Application Form (IRAS-Form-30102018)		30/10/2018
Letter of invitation to participant (Participant invitation letter)	3	08/10/2018
Letters of invitation to participant (invitation letter reminder)	1	24/09/2018
Non-validated questionnaire (Demographic questionnaire)	1	07/08/2018

Northern Health and Social Care Trust

www.northerntrust.hscni.net

Document	Version	Dated
Other (Response to PR Sub-Committee Queries)		20/11/2018
Participant consent form (Consent Form)	2	24/09/2018
Participant information sheet (PIS)	3	20/11/2018
Referee's report or other scientific critique report (filter committee report – RG3)		05/10/2018
Research protocol or project proposal (qualitative protocol)	3	02/10/2018
Summary, synopsis or diagram (flowchart) of protocol in non-technical language (Study flowchart)	2	22/08/2018
CV - CI Dr Alison Porter - Armstrong		Not dated
GCP - CI Dr Allison Porter-Armstrong		12/12/2018
CV - Mrs Patricia McIlwaine		October 2018
GCP - Mrs Patricia McIlwaine		15/03/2017
CV - Miss Beverley Turtle		13/10/2016
GCP - Miss Beverley Turtle		13/10/2016
CV - Dr May Stinson		Not dated
GCP- Dr May Stinson		16/02/2017
CV - Mrs Laurene Abbi		January 2007
GCP - Mrs Laurene Abbi		20/12/2018
CV - Mrs Jennifer Trainor		09/08/2018
GCP – Mrs Jennifer Trainor		11/01/2019

The following personnel have been approved to work on the study at this Trust:

Name	Indemnity Provided by
Dr Alison Porter -Armstrong	UU
Mrs Patricia McIlwaine	NHSCT
Mrs Laurene Abbi	NHSCT
Mrs Jennifer Trainor	NHSCT
Miss Beverley Turtle	UU
Dr May Stinson	UU

HSC Northern Health and Social Care Trust
www.northerntrust.hscni.net

Permission is granted subject to the attached conditions and I would ask you to please ensure that all members of the research team are familiar with these. Failure to abide by these conditions will invalidate permission and may result in the cessation of the research

I wish you every success with your project.

Yours sincerely,

Dr Desmond Rooney
Head of NHSCT R&D

CC Mrs Patricia McIlwaine, NHSCT
Mrs Laurene Abbi, NHSCT
Mrs Jenifer Trainor, NHSCT
Miss Beverley Turtle, UU
Dr May Stinson, UU
Mr Nick Curry, UU

Research & Development Office, Bush Road, Bush Road, Antrim. BT41 2QB
Telephone Number: 02894 424751

Conditions of Permission

Research Governance permission is issued provided the researcher(s) involved adhere to and abide by the conditions below.

- The researcher(s) must adhere strictly to the research protocol.
- There must be no changes to the research protocol or approved study documentation without the prior consent of the Trust, the Research Ethics Committee and, where applicable, the MHRA.
- Researchers must inform the NHSCT R&D Office if an extension to honorary contract is required for the duration of the study.
- There must be no changes in research staff without prior consent of the Trust.
- The Research Office should be informed if the Chief Investigator or Principal Investigator(CI/PI) is unable to continue to fulfil his/her duties as CI/PI for any reason such as long term absence, change in employment etc.
- There must be no increase in the resources required without prior consent of the Trust.
- Researcher(s) must report all untoward incidents and serious adverse events to the Trust.
- Any concerns in relation to the research protocol must be reported to the Trust.
- Researcher(s) must adhere to good research practice principles in line with the ICH Good Clinical Practice (GCP) guidelines.
- Researcher(s) must adhere to the Trust's Research & Development Standard Operating Procedures (available from the Research Office on request)
- On request, researcher(s) must make their research project available to Trust appointed monitors.
- The lead researcher must make an annual report to the Research Office for the duration of the project.
- The lead researcher should inform the Research Office on completion or termination of the project. Completion reports must be sent to the Research Office, Research Ethics Committee and, if applicable, MHRA.

Appendix 18.

Mirror therapy exercise protocol

Mirror Box Therapy Activity Booklet

Devised by

Occupational Therapy Rehabilitation Department
Whiteabbey Hospital

Northern Health and Social Care Trust

Mirror Box Therapy.

Mirror therapy is a specific therapy designed to strengthen arms and hands weakened by a stroke. In mirror therapy, we use movements of the stronger hand and arm to "trick our brain" into thinking that the weaker arm is moving. Researchers have shown that this "tricking of the brain" actually works – the brain areas responsibility for making the weaker arm move become stimulated.

You must keep looking the mirror whilst carrying out these activities.

Please complete the attached activities as recommended by your occupational therapist.

Occupational Therapist:

Whiteabbey Hospital

Activity 1 – Abduction/Adduction

Number of repetitions _____

Activity 2 – **Shoulder Flexion/Extension**

Number or repetitions _____

Activity 3 – Elbow Flexion/Extension

Number or repetitions _____

Activity 4 – Pronation/Supination

Number of repetitions _____

Activity 5 – **Wrist Extension**

Number of repetitions _____

Activity 6 – Finger Flexion/Extension

Number of repetitions _____

Activity 7 – Opposition Index/Middle Finger

Number of repetitions _____

Activity 8 – **Opposition Ring & Little Finger**

Number of repetitions _____

Appendix 19.

Participant invitation letter for Chapter 6

Occupational Therapy Department
Whiteabbey Hospital
Doagh Road
Newtownabbey
BT37 9RH
/ /2018

Dear PARTICIPANT NAME

Invitation to take part in study: Stroke survivors' views on mirror therapy in upper limb rehabilitation.

You may remember during your time in Whiteabbey Hospital using Mirror Box Therapy as part of your occupational therapy treatment. We would now like to invite you to take part in a research study where you will have the opportunity to give your views about the mirror therapy treatment. This study is being carried out to find out how best to deliver and support patient using mirror therapy in the future. It will be completed by researchers from Ulster University.

We have enclosed information sheet, please read it carefully and take time to think about whether or not you would like to take part. If you are interested in helping with this research, please complete the reply slip attached and return it in the stamped addressed envelope which is enclosed. If you agree to take part a researcher will be in contact with further details of the study.

Please be assured that the researchers have not been given any information about you and at this stage you are under no obligation to take part in the study.

If you would like more information before you make a decision, please contact the clinical lead occupational therapist, Mrs Patricia McIlwaine (028 9055 2326), who will be happy to answer any questions you have.

You will be free, of course, to withdraw from the study at any stage.

Yours sincerely,

Patricia McIlwaine/Lourene Abbi/Jennifer Trainor
Occupational Therapists Whiteabbey Hospital **Tel: 028 9055 2326**

Ulster University

Northern Health and Social Care Trust

Reply Slip

Your Name: _____

Address: _____

_____

_____

Telephone Number: _____

Yes, I am interested in hearing more about the study. I understand a researcher will contact me with more details and that I can change my mind about taking part at any stage.

Please tick here ☐

The best day(s) to contact me: Mon Tues Wed Thur Fri Sat (please circle)

The best time(s) to contact me:

Morning ☐

Afternoon ☐

Evening ☐

If you would prefer not to be contacted about the study, please tick the box below and return this reply slip to us in the stamped addressed envelope. Thank you for taking the time to read this letter.

☐

Appendix 20.

Participant Information Sheet for Chapter 6

Participant Information Sheet

Title of study: Stroke survivors' views on mirror therapy in upper limb rehabilitation

You have been given this information sheet because you are being invited to take part in a research study. This information sheet describes the study and explains what will be involved if you decide to take part. This study is being completed by Beverley Turtle, a PhD researcher based at Ulster University, together with the Occupational Therapy Department at Whiteabbey Hospital, Northern Health and Social Care Trust.

It is important for you to understand why the research is being performed and what participation would involve. Please take time to read the following information carefully and discuss it with others if you wish. You are welcome to contact us if there is anything that is unclear or for more information.

What is the purpose of the study?

This is a study designed to investigate experiences of those who have used mirror therapy during their stroke rehabilitation. You are invited to share your views and experiences of mirror therapy. This will help the research team to further enhance mirror therapy, to potentially benefit those with arm weakness after stroke.

Why have I been asked to take part in this study?

You have been invited to take part as you received mirror therapy as part of your treatment. As an expert user, your views and experiences will be used to further enhance therapy for arm recovery.

What will I have to do if I take part in this study?

You will be asked to attend a focus group discussion with individuals who have also experienced stroke. There are no right or wrong answers; we are interested in gathering your opinions. The discussion group will consist of between 6 - 8 people, lasting approximately 60 - 90 minutes.
This will take place at Whiteabbey Hospital. For travel to Whiteabbey Hospital, all travel costs will be reimbursed.
We ask your permission to audio record the conversation and to make the occasional note during the discussion, to help us remember the discussion accurately.

If you decide to take part, you may withdraw at any time without giving reason. This will not affect your medical care or legal rights.

What are the benefits of taking part in the study?

There are no direct benefits to taking part. However, you would be providing valuable benefits to the further development of mirror therapy, which could improve treatment for individuals who have had a stroke.

Is there any risk of taking part in this study?
There are no anticipated risks in taking part. However, taking part in a focus group may make you tired, refreshments will be provided throughout, you do not have to answer all questions and can take a break when needed.

Also talking about your experiences after stroke may be upsetting. You can skip any questions you do not wish to answer and can stop at any time, without giving reason. If the focus group upsets you, the research team will be able to assist you in making contact with appropriate supports, such as your G.P. or counselling service.

What happens if I withdraw from or am unable to continue with the study?

It is up to you if you decide to take part in the study. You are free to withdraw from the study at any time.

What will happen to any information I give?

If you take part in the study all personal information relating to yourself will be kept strictly confidential.
All information collected will be stored at Ulster University, Jordanstown Campus in a locked room and on a password protected computer, accessible only by the research team. All data will be stored in adherence with Data Protection (2018).
Results from the study may be submitted for publication in medical journals. Any extracts from what you say which are quoted in written work will be anonymised.

Contact details

If you would like any further details about this study or would like to ask any questions, then please do not hesitate to get in touch. I am the main contact:

Beverley Turtle
Email: turtle-b2@ulster.ac.uk
Tel: 07402804804
School of Health Sciences, Room 1F122, Ulster University. Jordanstown Campus.
Co. Antrim BT37 0QB

If you wish to contact a senior member within the university, or make a complaint, please contact:

Chief Investigator
Dr Alison Porter-Armstrong
Senior Lecturer in Rehabilitation Sciences at Ulster University
Email: a.porter@ulster.ac.uk
Tel: 02890 366651
School of Health Sciences, Room 1F120, Ulster University. Jordanstown Campus.
Co. Antrim BT37 0QB

You can alternately contact the Research Ethics and Governance Office at Ulster University and speak to or direct any complaints to:

Nick Curry
Head of Research Governance at Ulster University
Email: n.curry@ulster.ac.uk
Tel: 0289036 6629
Room 26A17, Research Office, Ulster University, Jordanstown, BT37 0QB

Or speak with contact the clinical lead occupational therapist in Whiteabbey:

Mrs Patricia McIlwaine
Clinical Lead Occupational Therapist, Northern Health and Social Care Trust
Tel: 02890552326
Occupational Therapy Department, Whiteabbey Hospital. Newtownabbey.
Co. Antrim. BT37 9HR

Thank you for taking the time to consider taking part in this study and we will respect your decision if you agree to take part or decline.

Appendix 21.

Participant follow up invitation letter for Chapter 6

Occupational Therapy Department
Whiteabbey Hospital
Doagh Road
Newtownabbey
BT37 9RH
/ /2018

Dear PARTICIPANT NAME

Invitation to take part in study: Stroke survivors' views on mirror therapy in upper limb rehabilitation.

Recently an invitation letter was sent to you, inviting you to participate in a research study. This follow-up letter is being sent to remind you to respond if you would like to take part. This study is being carried out to find out the views and experiences of stroke survivors who used mirror therapy as part of their stroke rehabilitation while at Whiteabbey Hospital. This study will be conducted by researchers from Ulster University.

Please read the enclosed information sheet carefully and take time to think about whether or not you would like to take part. Taking part in the study is voluntary. If you would like more information before you make a decision, please contact the clinical lead occupational therapist, Mrs Patricia McIlwaine, the Chief Investigator of the study (0289055 2326), who will be happy to answer any questions you have.

If you are interested in helping please complete the reply slip on the next page and return it in the stamped addressed envelope which is enclosed with this letter. A researcher who will be able to explain the study to you in more detail will then contact you. You will be free, of course, to withdraw from the study at any stage.
If we do not hear from you, we will assume you do not wish to take part and we will send you no further information.

Yours sincerely,

Patricia McIlwaine/Lourene Abbi/Jennifer Trainor
Occupational Therapists Whiteabbey Hospital **Tel: 028 9055 2326**

Reply Slip

Your Name: _____

Address: _____

_____

_____

Telephone Number: _____

Yes, I am interested in hearing more about the study. I understand a researcher will contact me with more details and that I can change my mind about taking part at any stage.

Please tick here ☐

The best day(s) to contact me: Mon Tues Wed Thur Fri Sat (please circle)

The best time(s) to contact me:

Morning ☐

Afternoon ☐

Evening ☐

If you would prefer not to be contacted about the study, please tick the box below and return this reply slip to us in the stamped addressed envelope. Thank you for taking the time to read this letter.

☐

Appendix 22.

Demographic information sheet for Chapter 6

Demographic information sheet

Participant pseudonym:

1. What is your age? _____

2. Gender:

Female ■ Male ■

3. Side of hemiplegia?

Right ■ Left ■

Appendix 23.

Blank consent form

Northern Health and Social Care Trust

Project title: Stroke survivors' views on mirror therapy in upper limb rehabilitation

Name of Chief Investigator: Dr Alison Porter-Armstrong

I confirm that I have read the information sheet for the above study. I have had the opportunity to consider the information, ask questions and had these answered satisfactorily. ☐

I understand that I am free to withdraw at any time without giving any reason. ☐

I agree to participate in the above focus group. ☐

I agree to have the focus group audio-recorded, and understand that this will then be transcribed. ☐

I understand that any personal data that could be used to identify me will be removed from the transcript of the focus group and that I will not be identified in any publications, reports or presentations. ☐

I agree to take part in the above study. ☐

I would like to receive a summary of study results. ☐

Name of participant (print)	Signature	Date
……	…….	……

Name of Researcher (print)	Signature	Date
……..	……..	……

Appendix 24.

Focus Group Topic Guide

Stroke survivors' views on mirror therapy in upper limb rehabilitation

Introduction

Welcome and Instruction

Thank you for agreeing to participate in a discussion about mirror therapy.

The aim of this focus group discussion is to find out your views and experiences of mirror therapy. You are invited because each of you received mirror therapy as part of your arm treatment following stroke. It is hoped the conversation from today can inform the future rehabilitation of others with arm weakness after stroke. The focus group discussion should take no more than 90 minutes. Feel free to help yourselves to refreshments throughout.

- We are tape-recording the conversation. However, I would like to assure you that the discussion will be kept confidential. The conversation is being recorded for us to transcribe and look at later. Only the research team will ever hear the tape, and nobody will be named on the transcript.

Ground rules

- We ask that only one person speaks at a time.
- You do not have to speak in any particular order
- As this is a focus group discussion, we would like you all to talk to each other; would like to get your views on what's important not ours; not testing you – no right or wrong answers.
- Everything that is discussed is confidential and we ask you not share this outside the room.
- Does anyone have any questions?

Ice-breaker

First, I'd like everyone to introduce themselves. Can you tell us your name?

Topic areas

1. What were your initial impressions of mirror therapy?
   Prompts:
   E.g. Can you describe how the mirror box works? What were you told?
   What were your expectations of mirror therapy?
   Were there any unanswered questions?
2. What did you think or feel about using mirror therapy as part of stroke rehabilitation?
   Prompts:
   E.g. What did you like about it?
   What did you find difficult about it?
   Did you believe in the illusion of your affected arm moving in the mirror?
   What were your thoughts on the amount of mirror therapy received during your stay in hospital?

3. What did you think about the movements you completed with the mirror box?
   Prompts:
   E.g. Were they easy to follow with the therapist?
   Did you attempt the movements with your arm that was in the box? How did you find that?

Did you feel able to complete the mirror box activities on your own, beyond your sessions with the occupational therapists? How so?
Did you use it on evenings/weekends, with family/friends or caregivers? Were instructions provided to family/ friends or caregivers?

4. What were your thoughts on the booklet provided?
   Prompts:
   E.g. What alternative formats do you think could be used?
   What other exercises would you want to see included, if any? Use of objects?

5. Regarding the design of the box itself, what were your thoughts on the mirror box device?
   Prompts:
   E.g. How easy was it to use?
   Did you need assistance to use it?
   What were your thoughts on the size of the mirror?
   Have you any recommendations to improve the mirror box design?

6. Once home, for those who took their boxes home, did you continue with mirror therapy?
   Prompts:
   E.g. Why did you continue using the mirror box at home?
   What helped you to continue with this, family/caregiver? Booklet?
   Were there any problems in continuing mirror therapy at home?
   Would you have any concerns about using mirror therapy at home?

7. Overall, have you any recommendations to improve the mirror therapy protocol?

8. Is there anything else we haven't touched upon, which you would like to discuss?

This is a semi-structured focus group topic guide, allowing for further exploration of participant experiences as they arise throughout the focus groups.

Appendix 25.

Extract from transcript

M: What factors help to keep you motivated?

P3: Well it's very difficult to try and tell somebody what, how to do it, its, its but it's very easy to convince yourself because I know I was able to do things before I left hospital, and I can't unfortunately do those now and I think the only reason for that is that I actually didn't feel any pain so I says right well leave it alone, it's not doing you any harm, its, the arms there. But to keep moving it because you need to keep the muscles moving. [LONG TERM PROCESS: Maintenance of self-directed therapy]

P2: Muscles depreciate very quickly.
M: Even the things that are easy just keep doing it.

P2: … Yep, yep.

M: Yep, mm-hmm.›

P1: I would have liked to have more time spent on my arm. Because I still have very little movement, movement in it. But for the walking it was great. And sorta once I could do one thing I figured I could do that what's next. Put that in the bag and move on. But I felt I needed more exercises on my arm, there wasn't enough of that [HEALTHCARE BARRIERS: Imbalance between upper and lower limb rehab focus]. Because I understand a BIT about the nerves trying to send signal down it, but you're right, if you don't use the muscles they waste away very, very quickly and that stiffness that comes, yes.

M: Regarding the design of the mirror box itself, what were your thoughts on the actual product [indicated mirror box on table]

P3: Well, M, I would have thought that the little corners that are cut off I would have thought if, particularly the one that's closest, the bit of the mirror that's closest to you, if that mirror, if the mirror came right down into that corner, because at times uh with me, I was having to turn the mirror in a position to see what my right hand was, was doing to, to make me think it was actually my left hand, because that piece of the mirror was missing and it was just restricting me being able to look into the mirror [PRACTICAL: increase mirror surface area on box].

Appendix 26.

Main RCT protocol

Methods/Design

Objective

This paper aims to describe the study protocol of a multi-centre randomised controlled trial of mirror box therapy (MBT) in upper limb rehabilitation with a sub-acute (0-3 months) stroke population. Specific objectives are to: 1) determine the upper limb movement, functional, quality of life and occupational gains acquired through mirror box therapy between baseline and hospital discharge and at 12 weeks post discharge; 2) explore the effect of time, treatments and patient differences on upper limb movement, functional, quality of life and occupational gains across all measurement points.

Design

The study is a multi-centre single-blinded randomised parallel group control trial. Participants will be randomly allocated to receive either standard care (standard occupational therapy upper limb rehabilitation) or standard care alongside mirror box therapy. Outcome measures will be recorded by the researcher blinded to group allocation at baseline, every 2 weeks as an in-patient, at discharge and 12 weeks post discharge.

Development of the study protocol

The content of the intervention and the outcome measures are based upon previous pilot work exploring the use of mirror box therapy with a sub-acute stroke population funded by the United Kingdom Occupational Therapy Research Foundation through a Research Priority Grant 2014. This study validated the overall approach with 40 sub-acute stroke participants, and the data from the study enabled the sample size for the current study to be calculated, and minor protocol adjustments made. These include:

1. more comprehensive assessment at the screening stage of participant recruitment including the addition of a Fatigue Scale (per Drummond *et al*, 2017), specific cognitive assessment to assess potential of acquisition of mental practice skills, and assessment of upper limb rehabilitation-potential using stage of recovery (Viatherapy, 2017);
2. adjustment to the sequencing of selected primary and secondary outcome measures;
3. measurement of outcome measures post-discharge upto 3 months only.

Participants and recruitment

Eligible participants will be hospital in-patients who have sustained a stroke within the past 0-3 months. We aim to recruit 180 patients to participate in the study from 4 Health and Social Care Trusts within Northern and Southern Ireland.

Screening of study population

Within 1 week of admission, the occupational therapist will screen each patient admitted to the hospital site for rehabilitation following stroke and identify all stroke patients who are: 1) deemed as requiring upper limb rehabilitation as a key component of their occupational

therapy treatment; 2) assessed as demonstrating readiness for upper limb rehabilitation based upon stage of recovery and time post stroke (Viatherapy, 2017); 3) assessed as not suffering from post-stroke fatigue using the Fatigue Severity Scale of the Fatigue Assessment Inventory (Schwartz, Jandorf & Krupp 1993); 4) assessed as not being cognitively impaired using the Montreal Cognitive Assessment (MoCA) (Nasreddine et al. 2005); 5) deemed to be able to understand a minimum of two part written or spoken commands in the English language. (See Figure 1).

Participants screened as meeting the above criteria will be flagged as potentially eligible for study inclusion to the Clinical Research Nurse (Northern Ireland sites) or the site-specific study gatekeeper (Southern Ireland sites) for recruitment and consent to the study.

Recruitment and consent

The clinical research nurse or site-specific study gatekeeper will meet with each patient and complete the screening process using the inclusion/exclusion criteria and provide a verbal explanation and a written information booklet to those who are eligible to participate. They will return to each patient 24 hours later and obtain written informed consent for participation. Once consent has been gained, the medical consultant will be informed that the patient has consented to take part in the study and the nurse or gatekeeper will ensure that all potential participants are fitted with an extra identity band to ensure there is one on each wrist. This is a vital requirement of the therapy to ensure the brain cannot identify that it is the non-affected arm reflected in the mirror that is being viewed rather than viewing the affected arm. It is also important that the control group wear the extra identity band to ensure blinding of the researcher acting as independent outcome assessor.

Allocation

Once a patient has given consent to participate, both the treating therapist and the researcher will be informed. The researcher will commence baseline assessment. The treating therapist will make direct contact with the guardian of the allocation (see Randomisation procedure below) to disclose whether the patient has been allocated to either standard care to standard care plus mirror box therapy. The therapist will then commence either ‘standard’ (control group) or ‘standard plus MBT’ (intervention group) upper limb rehabilitation depending upon group allocation. The researcher shall remain blinded to group allocation throughout the study. Recruitment of participants to the study will start in January 2019.

Study inclusion and exclusion criteria

To take part in the study, participants must: 1) be 18 years and over and newly admitted inpatient of the rehabilitation ward; 2) have a first diagnosis of CVA in the last three months resulting in upper limb motor loss; 3) be able to follow two part spoken or written commands in the English language; 4) have upper limb therapy designated as a main portion of goal directed treatment programme; 5) be able to perform at least one of the upper limb movements as per the Viatherapy application for post stroke arm recovery (Viatherapy, 2017); 6) score 35 or below on the Fatigue Severity Scale of the Fatigue Assessment Inventory; 7) score above 19 on the Montreal Cognitive Assessment, and, 8) consent to take part in the study. Participants will be excluded if they do not meet the above criteria.

Ramdomisation and concealment
Block randomization within each site will be undertaken by the statistician using a computer- generated randomization list. Block randomization will enable a balance of allocation to each study arm within each site, as well as across the four trial sites. The statistician shall prepare randomization lists for each site and allocation will be concealed in consecutively numbered, opaque sealed envelopes. An individual independent of the study will act as guardian of the envelopes to ensure there is no bias in the allocation and will disclose allocation on a sequential basis to the treating therapist.

Sample size
Data from the pilot study were used to calculate the sample size necessary for this study. The graded Wolf Motor Function Test scores were used to determine the sample size. To detect a clinically meaningful difference of 0.7 (SD 1.45) between the two groups at post-intervention on the function score, 90 participants per group (180 participants in total) are required to achieve power of 90% at an alpha of 5% (2-tailed).

Intervention
Control Group
Participants shall receive their standard occupational therapy treatment for upper limb rehabilitation for the duration of their in-patient stay, which is 3-5 sessions per week of approximately 45 minutes duration. This classic rehabilitation treatment is based upon neurodevelopmental theory using the Bobath approach of postural control and repetitive task training. This will follow the procedure used within the pilot study (Northern Health and Social Care Trust 2018) and progresses through 8 phases from assisted to unassisted movements, gross upper limb movements to wrist and fine finger movement, using remedial and functional activities as well as ward-level rehabilitation.

Intervention Group
In addition to the standard occupational therapy treatment outlined above, participants allocated to the intervention group will be required to perform two 20-minute sessions of mirror box therapy, five days/week for the duration of their in-patient stay. Also based upon neurodevelopmental theory, this treatment creates the illusion of perfect bilateral synchronization (Oujamaa et al 2009) of repetitive task training by concealing the affected arm in a mirrored box that reflects the repetitive upper arm movements conducted by the unaffected limb. Sessions will be conducted at the patient's bedside or in the occupational therapy department. Participants will be seated in a comfortable high chair and positioned in front of an adjustable height table. The mirror box will be positioned on the table in front of the participant. The participant will place or be assisted by the therapist to place the affected arm into the open-end of the nylon box; the mirror section will face the patient's non-affected side. The participant will then be instructed to look into the mirror whilst completing the programme of eight gross and fine motor movements.

Outcome Measures

Primary Outcome Measure

Graded Wolf Motor Function Test (gWMFT) (Morris et al, 2001): The gWMFT is a 15-item standardised measure which determines the motor ability of participants by recording functional movement time (0- 120 seconds per item, total = mean of 15 items, maximum score= 120 seconds) and quality of movement (0-7 Likert scale per item with 0= no movement, to 7 = normal movement, total = mean of 15 items, maximum score= 7). This graded version has two levels of each task which can be chosen depending on the participants' general functioning level. The graded version was developed from the original WMFT which has shown to have good reliability and validity (Morris et al, 2001). Despite being named a motor assessment, this assessment includes assessment of the upper limb using functional activities and, as such, is considered of relevance to OT outcomes. This outcome measure has also been used by other investigators in previous studies with a stroke cohort (Uswatte et al, 2005) and has shown to have good inter- and intrarater reliability for performance time and functional ability in the recent pilot study by the authors (Turtle et al, 2017).

Secondary Outcome Measures

Functional Independence Measure (Uniform Data System for Medical Rehabilitation 1996): The FIM is an 18-item measure of 6 areas of function (self-care; sphincter control; mobility; locomotion; communication and social cognition) grouped into two domains of motor items and cognitive items. Each item is scored on a 7-point likert scale and the score indicates the amount of assistance required to perform each item (ranging from 1 representing total assistance in all areas to 7 representing total independence in all areas), and has been widely used post-stroke (Granger et al 1993; Beninato et al, 2006).

EQ-5D-5L (Oemar and Janssen, 2013; Van Reenen and Janssen 2015): The EQ-5D-5L is a widely-used standardized 2 page instrument for use as a measure of health outcome. It is applicable to a wide range of health conditions and treatments and provides a simple descriptive profile of 5 domains (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) across 5 levels ranging from no problems to extreme problems, and a visual analogue scale of the respondent's self- rated health ranging from 'best imaginable health state' to 'worst imaginable health state'. Normative scores exist for the UK population against which outcomes can be compared (Kind et al 1999).

Canadian Occupational Performance Measure (COPM) (Law et al, 2005): The COPM is a standardized outcome measure to detect change in a client's self-perception of occupational performance over time. It uses a semi-structured interview format and structured scoring method to detect change scores between assessment and reassessment in everyday occupational activities.

Other Measure

Self-report exit questionnaire*:* This questionnaire aims to gain insight into the subjective opinion and establish participant impression and understanding of MBT and provide an opportunity for participants to provide feedback to the therapy team. On completion, all trial participants will be asked to complete a self-report exit questionnaire to gain insight into the subjective opinion and establish participant impression and understanding of upper limb rehabilitation, occupational therapy intervention and mirror box therapy and provide an opportunity for participants to provide feedback to the therapy team.

Data Analysis

To determine the upper limb movement gains acquired through mirror box therapy between baseline and hospital discharge using the graded Wolf Motor Function Test, differences from baseline to discharge will be analysed using ANCOVA, with baseline assessment as the covariate. This same analysis will be applied to answer the secondary objectives of function and quality of life using the outcomes from the Functional Independence Measure and EQ-5D-5L.

To determine the sustainability of upper limb movement gains acquired through mirror box therapy at 12 weeks post discharge using the graded Wolf Motor Function Test, differences from discharge to 12 weeks post discharge will be analysed using ANCOVA with discharge assessment as the covariate. This same analysis will be applied to answer the secondary objectives of function and quality of life using the outcomes from the Functional Independence Measure and EQ-5D-5L.

To explore the effect of time, treatments and patient differences on upper limb movement across all measurement points using the graded Wolf Motor Function Test, a linear mixed model will be used with all measurements across time. This same linear mixed model will also be used to explore the effect of time, treatment and patient differences on the secondary outcomes of the Functional Independence Measure and EQ-5D-5L.

Qualitative analysis, using a thematic approach, will be used to explore the change in occupational performance goals for both arms from baseline to discharge and discharge to 12 weeks post discharge using the Canadian Occupational Performance Measure.

Timeframe for study

The study will commence on 1st January 2019 for a period of 36 months.